AD-A265 070

DTIC
ELECTE
MAY 27 1993
S C D

Final Tile Design Study
for the
Digital Chart of the World

Contract No.: DMA600-89-C-0023
CDRL B001

Date Prepared: November 1990

Prepared for:
Defense Mapping Agency Systems Center
8613 Lee Highway
Fairfax, Virginia 22031-2138

Attn: Ms. Lynn Martin
Contract Specialist

Prepared by:
Environmental Systems Research Institute, Inc.
380 New York Street
Redlands, California 92373
U.S.A.

CONTENTS

November 1990 Final Study, July 90 - November 90

Final Tile Design Study for the Digital Chart of the World

C: DMA600-89-C-0023

Environmental Systems Research Institute, Inc.

Environmental Systems Research Institute, Inc.
380 New York Street
Redlands, CA 92373

Defense Mapping Agency
8613 Lee Highway
Fairfax, VA 22031-2137

DMA/TR--93/03

Distribution Unlimited (A)

This study completes earlier investigations and makes final recommendations for the partitioning of large spatial databases undertaken in the development of the Defense Mapping Agency's new vector product, the Digital Chart of the World (DCW). The DCW is a geo-spatial, topologically structured, global database designed for use in Geographic Information Systems (GIS). The study covers the characteristics of tiling, or partitioning, of large digital geo-spatial databases to improve data storage, access, and use. It finalizes tiling requirements, goals, and studies and makes final recommendations for the tiling system used in the DCW. It is the second of two tiling reports performed under the DCW development effort. The first was the Interim Tile Design Study for the Digital Chart of the World, July 1990.

Geo-spatial database partitioning, tiling, Geographic Information System, GIS, spatial databases, Vector Product Format, VPF, Digital Geographic Information Exchange Standard, DIGEST, Geographic Information Standards, digital cartography, digitizing, Geo-spatial Information, CD-ROM performance, Defense Mapping Agency, DMA, Vector Maps, Digital Chart of the World, DCW.

42

Unclassified Unclassified Unclassified UL

1.0 INTRODUCTION

This document compares the merits of fixed and adaptive tiling for the Digital Chart of the World (DCW), a general-purpose global geographic database distributed on Compact Disc-Read Only Memory (CD-ROM) that is derived from the 1:1,000,000-scale Operational Navigational Chart (ONC) series and accompanied by display software designed for Personal Computers (PCs). The findings to date indicate that a fixed tiling scheme is a more practical solution than an adaptive tiling scheme in view of the constraints under which the database is being produced and will be maintained and used.

1.1 PURPOSE

This paper, the Final Tile Design Study, presents both the results of work completed since the Interim Tile Design Study and our final recommendation for the DCW tiling scheme.

1.2 BACKGROUND

Data density varies greatly from one ONC sheet to another, and since the ONCs are the source for the database, it is important to consider this variation in selecting a tiling method for the DCW. Because of its ability to effectively manage large variations in data density, an adaptive, rather than a fixed, tiling scheme was believed to be the best candidate at first. It was believed that a performance improvement could be obtained through the use of regular, systematic quadtree partitions. However, as the DCW evolved through a series of prototypes it became clear that the value of managing variations in density would have to be weighed against the disadvantages of any adaptive tiling scheme for other aspects of the DCW project, particularly production.

By the time Prototype 3 was released, the recommendation for the tiling scheme had shifted from adaptive to fixed. The reason for the shift was explained at the Project Detail Design Review in August. By that time the effects of adaptive tiling on the project were understood well enough for adaptive tiling to be judged incompatible with the DCW. The recommendation was therefore made for the tiling to be fixed, and, further, for the fixed tiling to be regular. The recommended size of the fixed tile will be empirically determined after other, maturing elements of Prototype 4 can be evaluated.

1.3 THE NEED FOR TILING

The DCW database is composed of a set of files, each of which represents one component of the spatial and attribute content present. These simple files are assembled into an intermediate set of database structures called layers. The files required to represent layers are set by data structure and feature topology and for that reason differ in both type and number. For example, a 1000-record face layer will contain more files and be larger than a 1000-edge layer because faces require a more extensive structure than edges. Layers that share a common topology, such as Roads, Utilities, and Railroads, all of which are edges, will have identical representation in the database, differing only in the number of feature occurrences among the layers.

Both files and layers represent what are called vertical aggregations of data. Neither, however, expresses the horizontal component present in spatial databases (that is, the distribution of features over space). Because of physical constraints, such as the availability of Random Access Memory (RAM) or software limits, spatial databases greater than a certain size become too large to manage as a single unit and must be partitioned into smaller spatial units. The mechanism for breaking a large spatial database into smaller units is termed tiling, or partitioning, and results in units of manageable size.

1.4 REQUIREMENTS

Tiling is a management tool with no other function than making spatial datasets smaller. Geographic databases are tiled in order to make it possible to manage spatial data within a set of operational constraints, including the application, method of maintenance, data structure, software functionality, and characteristics of secondary storage devices. If there were no operational constraints on the database, there would be no reason to tile. However, since the current size estimate for the DCW exceeds the capacity of a single CD-ROM (the designated secondary storage device), the database will need to be tiled. In situations where one particular constraint dominates, tiling may be selected to optimize for that factor. However, a general-purpose database such as the DCW has no special dominating constraint and must, therefore, utilize a tiling scheme that is satisfactory from the standpoint of all constraints on the database.

This study identifies six requirements for tiling the DCW. First, the selected scheme must be global. Tiling schemes that are not global, such as that used for the Universal Transverse Mercator projection or that represented by the ONC map boundaries, are not acceptable for the database. Second, it must be possible to implement the scheme effectively. That is, the tiling scheme must be an integral part of the database development sequence and not adversely impact that process. Third, the tiling scheme must conform to Vector Product Format (VPF). Since any tiling scheme will be organized and managed as a VPF library, compliance with the requirements of this data format is mandatory. Fourth, the tiling scheme must be usable with both existing and future products. Fifth, only whole tiles may be placed on a single CD-ROM. Finally, sixth, the tiling scheme must be compatible with the indexing scheme used for the DCW (which is discussed in the Final Indexing Studies, CDRLs B003–B006). These requirements, addressed together, will yield a tiling scheme that creates manageable database units.

Figure 1 illustrates the DCW production process. Map sheets are automated first; they are tiled afterwards by the production software. Then the data are converted to the format used for the DCW (VPF); and, finally, the data are transferred to the CD.

1.5 TILING TYPES

There are two general approaches to spatial partitioning. The first fixes the amount of data that any tile may contain by varying the area within which it is held. This approach is commonly referred to as adaptive tiling and is, by definition, data driven. The second method sets the physical area of the partition and allows the volume of data within it to vary. This method is commonly referred to as fixed tiling. One of the main objectives of partitioning Prototype 3 into a number of different tiling schemes was to evaluate the relative merits of fixed and adaptive tiling for the DCW.

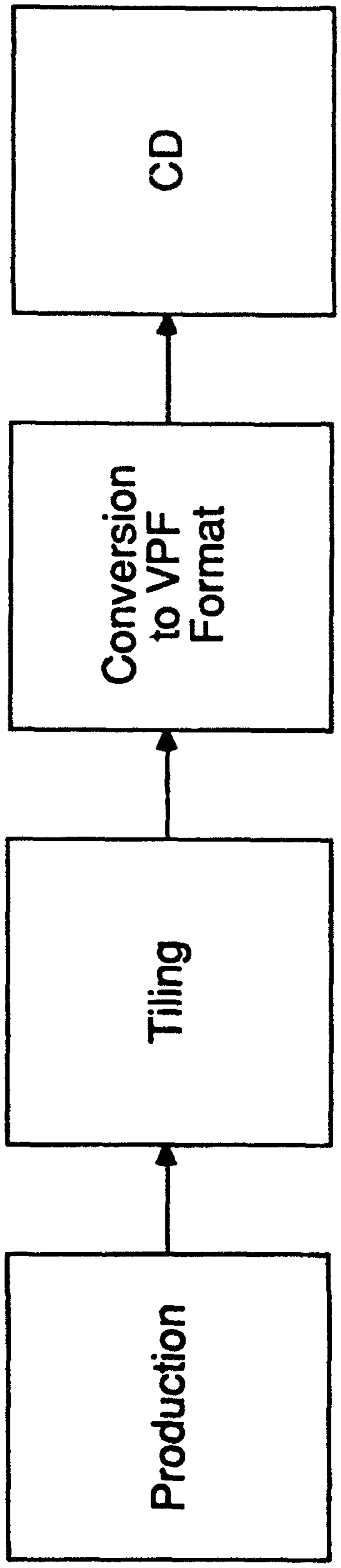

Figure 1. DCW Production Process.

2.0 ADAPTIVE TILING

Spatial databases that exhibit high feature heterogeneity are candidates for adaptive tiling. Adaptive tiling is a procedure that generates a set of variably sized spatial partitions, no member of which may contain more than a given volume of data. Adaptive tiling responds directly to any high level of feature heterogeneity. Table 1 presents the measured data volumes in bytes for the map sheets produced for Prototype 3. This table illustrates the high feature heterogeneity present in the DCW.

Table 1. Byte Totals for ONCs E-18, G-18, and N-13 and Jet Navigation Chart (JNC) 120.

Sheet	Total Bytes
ONC E-18	15,439,402
ONC G-18	7,617,531
ONC N-13	1,677,379
JNC 120	1,350,492

Figure 2 is a modeled surface showing data density distribution for the DCW using the ESRI production estimates of ONC/JNC feature counts as a source. The surface is expressed in units of standard deviation from a mean data density of 0. Increasing darkness indicates increasing volume. Therefore, if adaptive tiling were used, where the deviation from the mean is greater than the average, tiles would be expected to increase in number while decreasing in spatial extent. For the DCW, adaptive tiles would be smallest in the Himalaya Mountains and central Canada. Conversely, where data density is less than the average, tiles should be larger in spatial extent and smaller in number.

In order to effectively implement an adaptive tiling scheme, a procedure must be developed to satisfy two basic requirements. First, a procedure must be developed to automate the spatial partitioning process; and second, both a value and a unit of measure for data volume must be determined. The objective of the partitioning process would be to produce tiles that contain no more than this target data volume; and, as such, the process would be highly sensitive to data content.

2.1 ADAPTIVE TILING PROCEDURE

An adaptive tiling procedure is illustrated by the example described in Sections 2.1.1 through 2.1.6. Assume that the source maps proceed through a production sequence that does not allow the entire database to be completed before tiling occurs; sheets will need to be tiled as they are prepared. Also assume that the tiling scheme is defined on a plane longitude/latitude grid system with its origin at the lower left corner (180 degrees West and 90 degrees South). Again, assume that the adaptive tiling scheme to be used is a systematic spatial quartering and that the four tiles produced with each quartering are numbered clockwise from 1 to 4, starting with the upper left quadrant.

For purposes of this example, the tiling will have as its objective the coverage of the area of ONC map sheet F-1 (which happens to be the first map sheet produced for the DCW). ONC F-1 covers the area from 13 degrees West to 3 degrees East and from 40 degrees North to 48 degrees North. The relative position of ONC F-1 within the longitude/latitude grid system is shown in Figure 3. The next assumption is that ONC F-1, which covers 16 degrees of longitude and 8 degrees of latitude, contains 24 megabytes (MB) of data, or approximately 187,500 bytes per square degree. Finally, assume that the map sheet has been prepared for tiling and subsequent conversion. This requires that all scanning, vectorization, construction of topology, attribution, editing, and quality control checks have been completed for all the data contained within the map sheet.

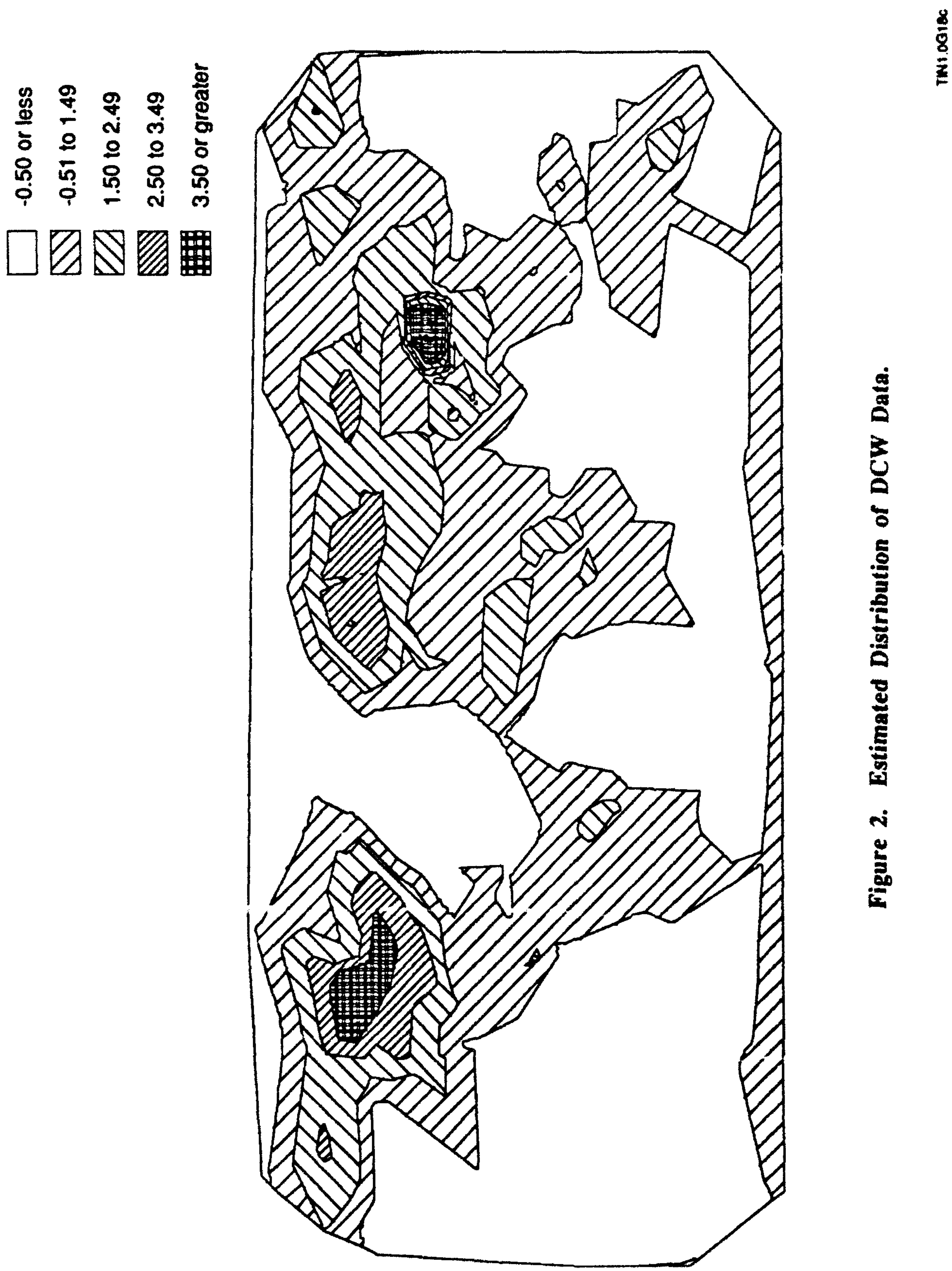

Figure 2. Estimated Distribution of DCW Data.

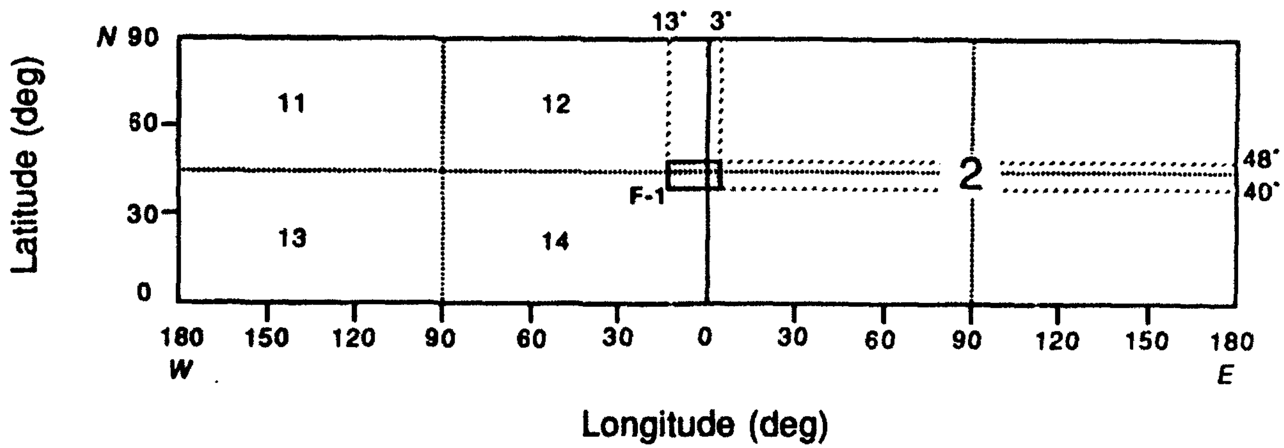

Figure 3. Level 2 Partitioning of ONC F-1.

Fig. JJ(c) Map Sheet

2.1.1 Step One: Determine If Partitioning Is Required

Given the conditions described above, the following is a procedure to implement systematic adaptive tiling. Compare the volume of data in the map sheet to the maximum value allowed for any tile. If the data volume in the sheet is greater than the maximum, the sheet will require partitioning. If the data volume is less than the maximum, the sheet must be spatially joined, or aggregated, with neighboring sheets before data volume is re-evaluated. If the adjacent map sheets are not available, then ONC F-1 cannot be completely tiled.

For this example, we will assume that partitioning will be required.

2.1.2 Step Two: Partition the Globe

Partition the grid system into tiles until the first occurrence of a tile boundary/map sheet intersection is found. Figure 3 shows that this happens to occur at the first global division, which divides the Earth along the equator and prime meridian. The prime meridian forms the boundary between tiles 1 and 2 and divides the map sheet into two sections of different size. The western section extends from 13 degrees West to the prime meridian and contains approximately 19.5 MB of data; the eastern section, extending from 3 degrees East to the prime meridian, contains approximately 4.5 MB of data.

2.1.3 Step Three: Estimate Data Volume

The data volumes for each of these two map sections produced by partitioning must be evaluated. If the data volume for a map section is less than the stated maximum, additional map data must be appended before a tile extent can be evaluated. If, on the other hand, the volume of data for the section produced by partitioning is greater than the stated maximum, that map section must be partitioned again. Assume, for purposes of this example, that the map section to the east, in tile 2, is below the allowable maximum and that the map section to the west of the prime meridian, in tile 1, is larger than the data volume maximum.

2.1.4 Step Four: Continue Partitioning and Estimating Data Volume

Consider the western map section that exceeds the allowable maximum (Figure 3). Since the data volume in the western section exceeds the maximum, tile 1, within which the western section lies, must be partitioned until the next tile boundary/map section intersection occurs. Figure 3 illustrates the result, which generates four tiles: 11, 12, 13, and 14. The level 2 division further splits the western map section, resulting from the first intersection, into two subsections. The volume of data in the northern portion, tile 12, is approximately 7.3 MB; the lower portion, tile 14, contains approximately 12.2 MB of data. If either of these map sections were to contain a data volume greater than the allowable maximum, continued partitioning would be required. Figure 4 illustrates the systematic patitioning of tiles 12 and 14 to level 4. Figure 5, which is focused on the area covered by ONC F-1, illustrates partitioning to level 5. Two tiles, 12443 and 12444, are produced from tile 1244. Tile 12443 contains approximately 1.0 MB of data, and tile 12444 contains approximately 6.3 MB of data. Level 5 partitioning of tile 1422 also yields two tiles: 14221, which contains approximately 1.6 MB of data, and 14222, which contains approximately 10.5 MB of data. Assume that at this level the map sections in tiles 14222 and 12444 still exceed the maximum. The western portion of Figure 6 illustrates partitioning to level 6 and yields the first occurrence of tiles that are completely filled with data (tiles 124443, 124444, 142221, and 142222). Table 2 lists the map sections and data volumes that result from partitioning the western portion of ONC F-1 to level 6.

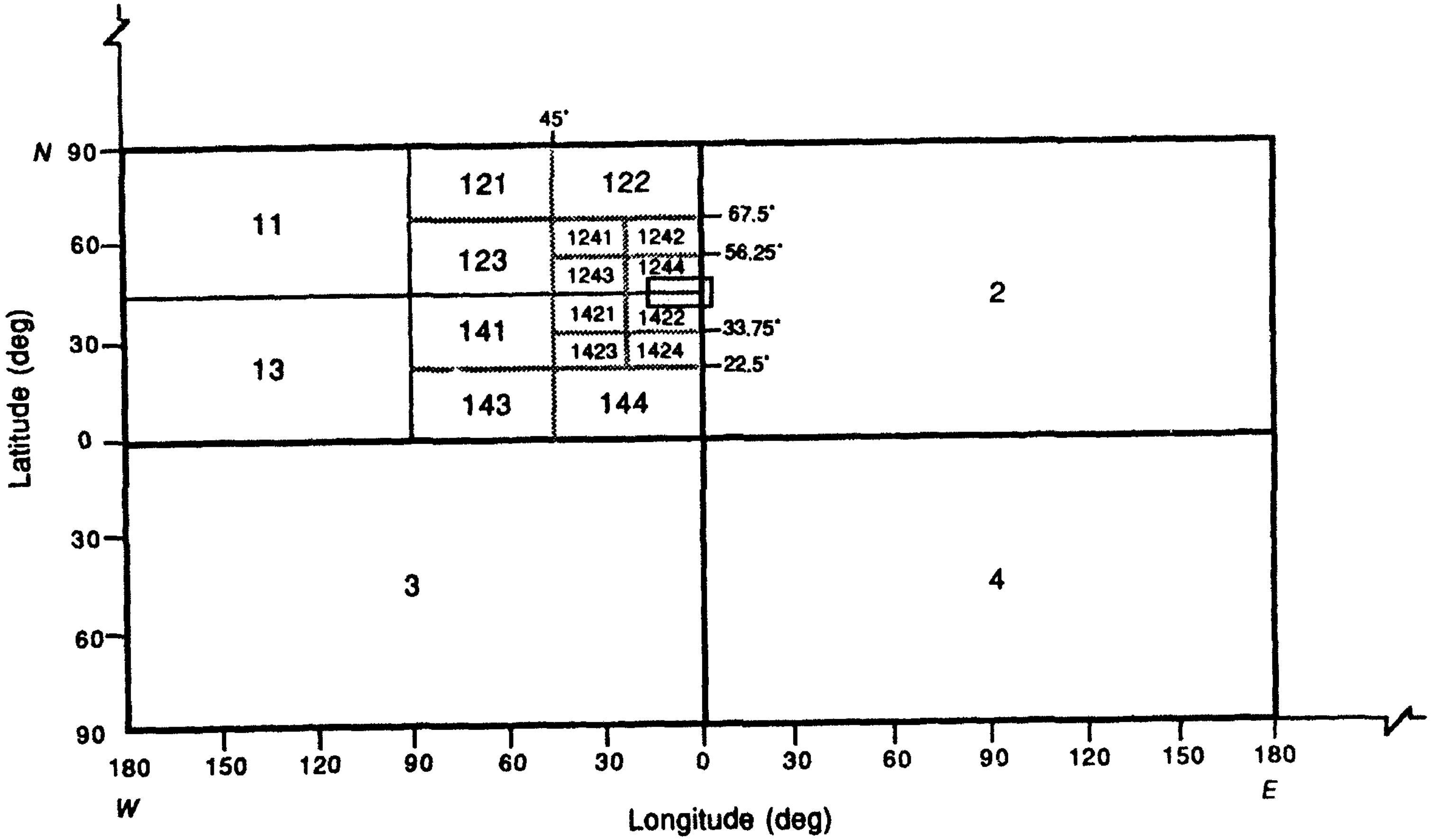

Figure 4. Level 4 Partitioning of ONC F-1.

Level 4 Partition.

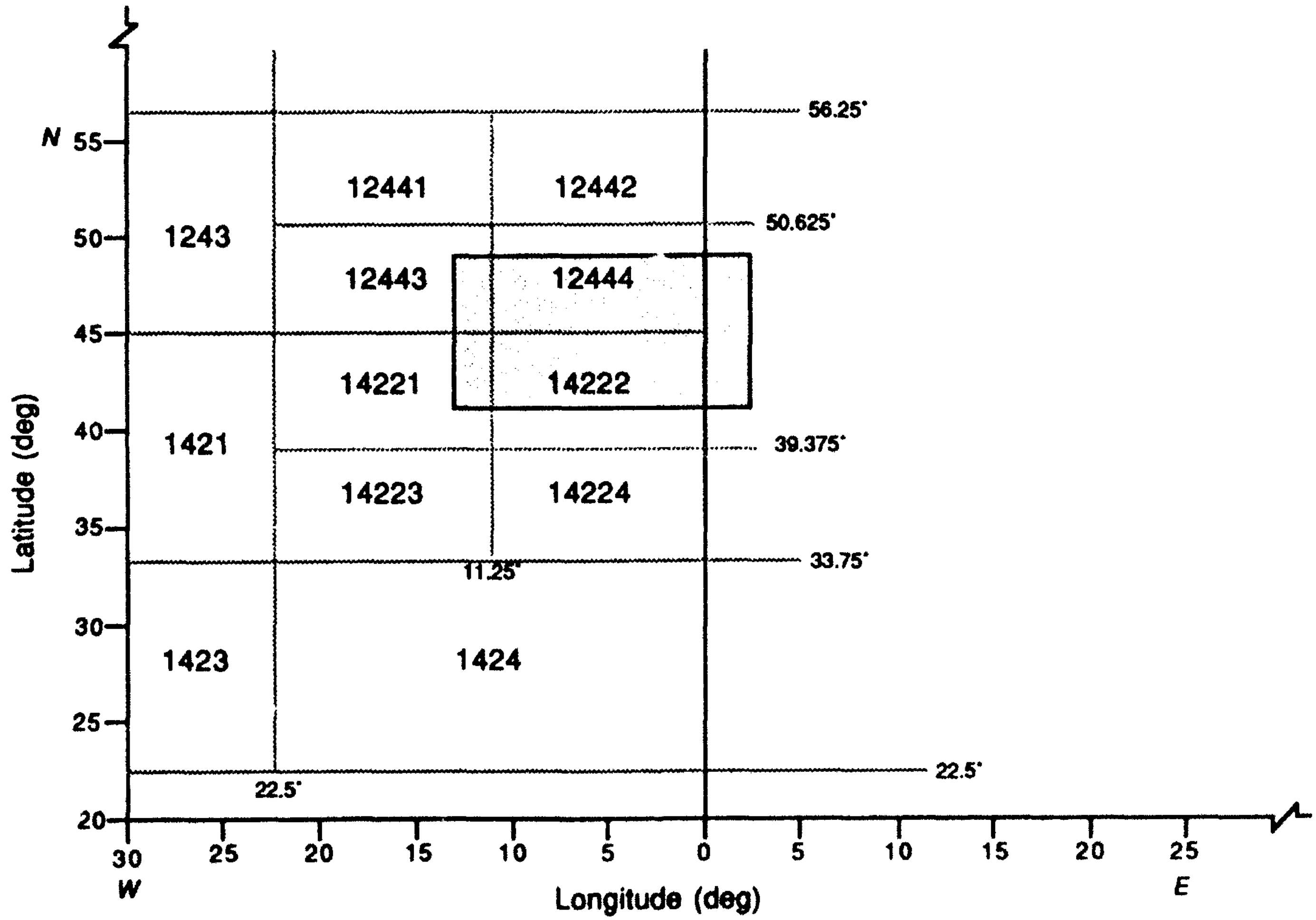

Figure 5. Level 5 Partitioning of ONC F-1.

ONC F-1 and Level 5

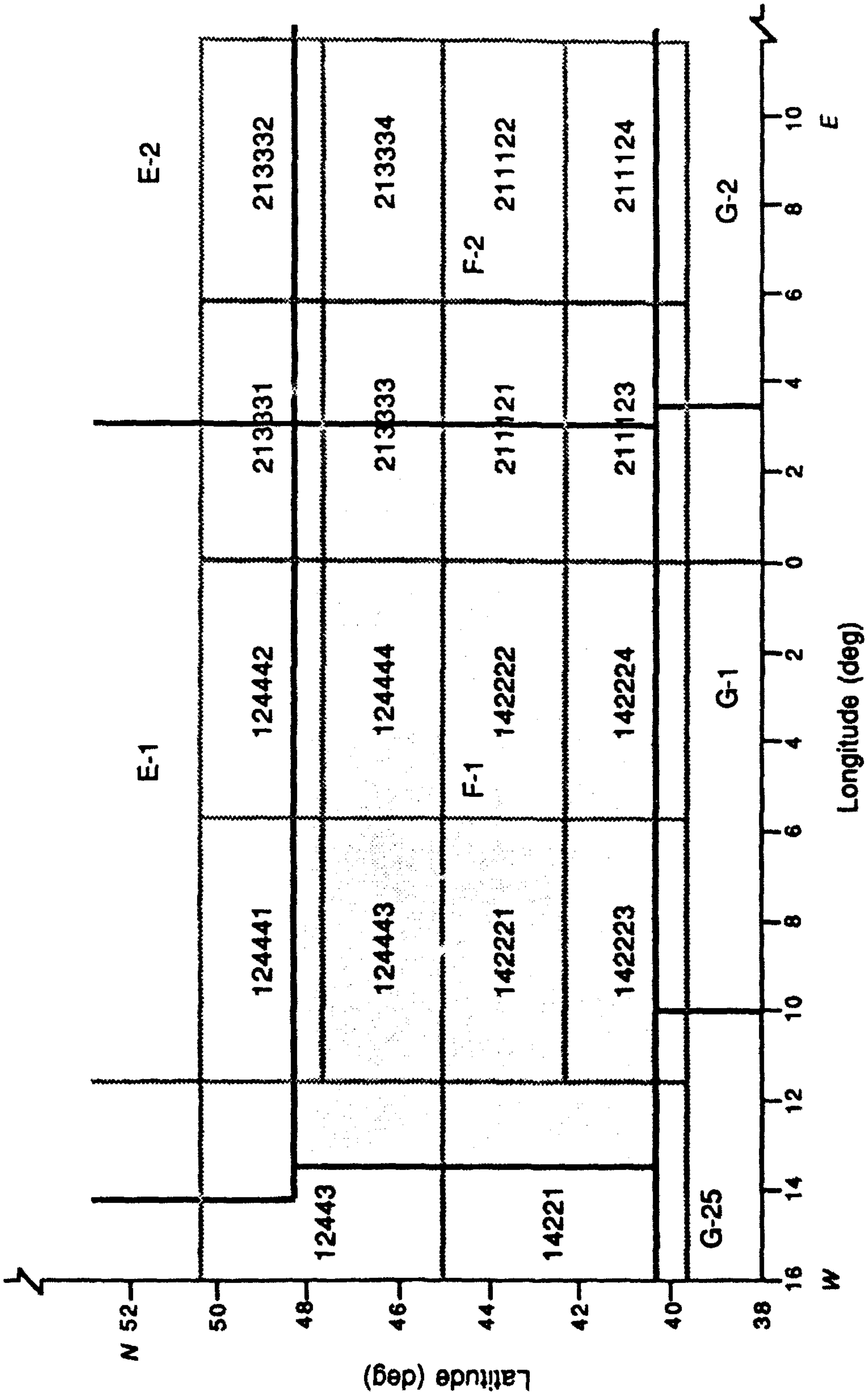

Figure 6. Level 6 Partitioning of ONC F-1.

Table 2. Data Volumes Resulting from Adaptive Partitioning of ONC F-1.

Partition	Area (percent)	Data Volume (MB)
Whole Sheet	100	24.0
Level 1		
Tile 1	81	19.5
Tile 2	19	4.5
Level 2, Tile 2	19	4.5
Tile 12	30	7.3
Tile 14	51	12.2
Level 5, Tile 2	19	4.5
Tile 12443	4	1.0
Tile 12444	26	6.3
Tile 14221	7	1.6
Tile 14222	44	10.5
Level 6, Tile 2	19	4.5
Tile 12443	4	1.0
Tile 14221	7	1.6
Tile 124441	1	0.2
Tile 124442	1	0.2
Tile 124443	12	3.0
Tile 124444	12	3.0
Tile 142221	12	3.0
Tile 142222	12	3.0
Tile 142223	10	2.4
Tile 142224	10	2.4

2.1.5 Further Steps: Assemble with Adjacent Sheets

The four tiles 124443, 124444, 142221, and 142222, which at level 6 partitioning were completely filled with data, can be evaluated without any adjacent map sheet data. However, the seven map sections 2, 12443, 14221, 124441, 124442, 142223, and 142224, which are contained in tiles not filled with data, will require that additional map data be joined to them prior to another evaluation of tile content. Figure 7 identifies the map sheets needed to complete the geographic extents defined by incomplete tiles generated down to level 6 partitions. The areas contained in the tiling scheme that are west of ONCs F-1 and E-1 do not have ONC map sheet coverage and therefore contain zero data volume. Completing the western portion of ONC F-1 may require an additional partitioning of tiles 12443 and 14221, depending upon the amount of data added from ONCs E-1 and G-25, respectively.

With adjacent sheets available, tile 2 at level 1, which contains the eastern portion of the map sheet, must now be partitioned until the smallest nonintersecting tile that wholly contains the eastern map section is reached. Then adjacent map sheet data must be added in order to completely fill the tile defined. As shown by the eastern portion of Figure 6, this procedure is a mirror image of the tile 1 partitioning discussed earlier. Completing the eastern portion of the map sheet will require the joining of ONCs E-2, F-2, G-1, G-2, and the evaluation of the assemblage partitioned at level 6.

Adjacent Sheets to F-1

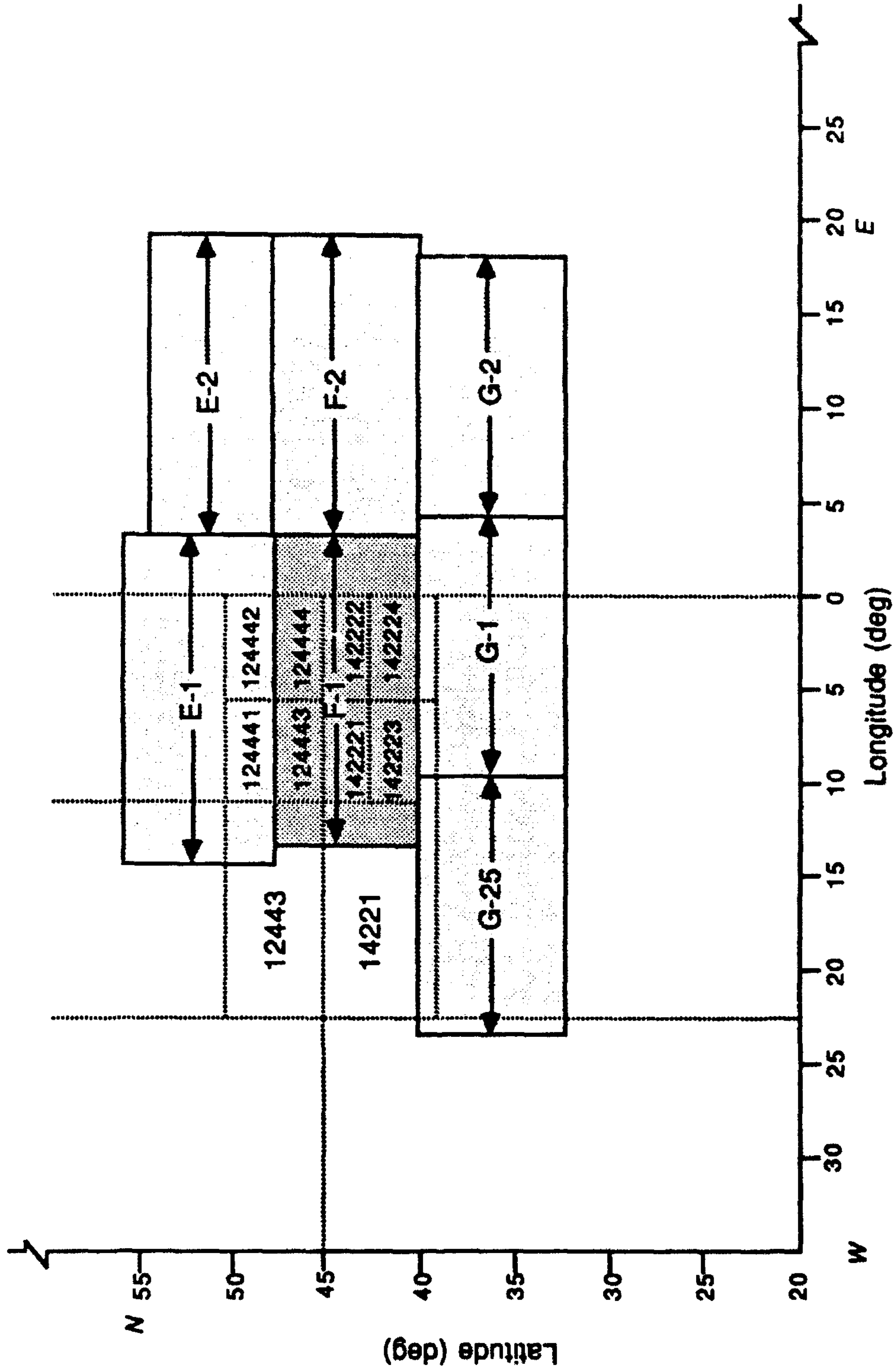

Figure 7. ONC Sheets Required to Complete Level 6 Tiling of ONC F-1.

The tiles needed to complete the eastern section of ONC F-1 are 213331, 213333, 211121, and 211123. The data volume of each incomplete map section that remains may now be estimated; tiles with too large a data volume can continue to be partitioned, and tiles with a data volume that is either too small or incomplete can be aggregated. New, incomplete map sections produced on the perimeter of the combined group of sheets will be identified as the partitioning process continues, as will the new ONCs required to spatially complete each tile.

2.1.6 Summary of the Adaptive Tiling Procedure

Figure 8 illustrates some characteristics of the tiling scheme that resulted from partitioning the first production map sheet (ONC F-1). First, the set of tiles identified with the gray tint has been correctly evaluated such that no one exceeds a given amount of data. However, the tiles may not necessarily contain equal amounts of data, since intra-sheet data density can also be high, as illustrated in Figure 9. This figure lists the distribution of data for ONC G-18 as a whole and for each of its four quarters. The reduced number of total bytes in the lower left quadrant is a result of the map sheet extending into the Pacific Ocean, which is an area with no data. A sharp line separates areas with data from areas with none (areas where selected map sheet data is not present or no ONC coverage exists). Also, the sum of the four quadrant byte values is greater than the whole sheet total, indicating that there is some amount of storage associated with tiling.

Second, to complete the tiling procedure, adjacent map sheets which have a portion of their extent within an incomplete tile must be automated in order to provide an evaluation of data volume and, therefore, to determine the appropriate level of tiling. In this example, 10 map sections, the tile numbers for which are contained within the boxes in Figure 8, will remain to be joined to adjacent map sheets. As can be seen by this example, an adaptive tiling procedure affects the production process by requiring that certain map sheets be prepared to continue tiling systematically from a given start sheet.

2.1.7 Nesting Requirements

The ONC F-1 example did not illustrate the implications of the absolute level of nesting required to properly partition the data. Since the distribution of all ONC data has only been estimated, the deepest level of nesting, or partitioning, required will not be known with certainty until the highest-volume map sheet is prepared and the tiling scheme is tested.

From the estimation procedure used in the preparation of Figure 2, ONC G-8 is believed to contain the largest data volume, with 89.5 MB of data. This sheet, although it has 2 degrees less longitudinal extent than ONC F-1, still has an average density of 799,000 bytes per square degree, which exceeds that of ONC F-1 by a factor greater than four. If it is presumed that ONC G-8 must continue to be partitioned until the data volume of any tile does not exceed the data volume stated as maximum for the example partitioning of ONC F-1, level 7 partitioning would be required for ONC G-8. At level 7, the data volume would be 3.2 MB; at level 8, it would be 588,000 bytes. The smallest data volume for a whole tile from the ONC F-1 example was 3.2 MB, which suggests that level 8 partitioning may be necessary.

Partitioning at levels of 8 and greater introduces an accuracy problem, however, because the tile boundary coordinates required to partition map sheets into smaller data volumes cannot be represented in single precision. The latitudes of the level 7 partitioning of ONC G-8 are from 37.96875 to 39.375, which are within single precision accuracy. Level 8 partitions for ONC G-8 are from 38.671875 degrees North to 39.375 degrees North. The value 38.671875 cannot be represented in single precision and will, therefore, be rounded to 38.67188. Although a requirement to partition to a level 8 division is contingent upon measured data volume, if it becomes necessary, coordinate generalization of the tile boundaries will be introduced.

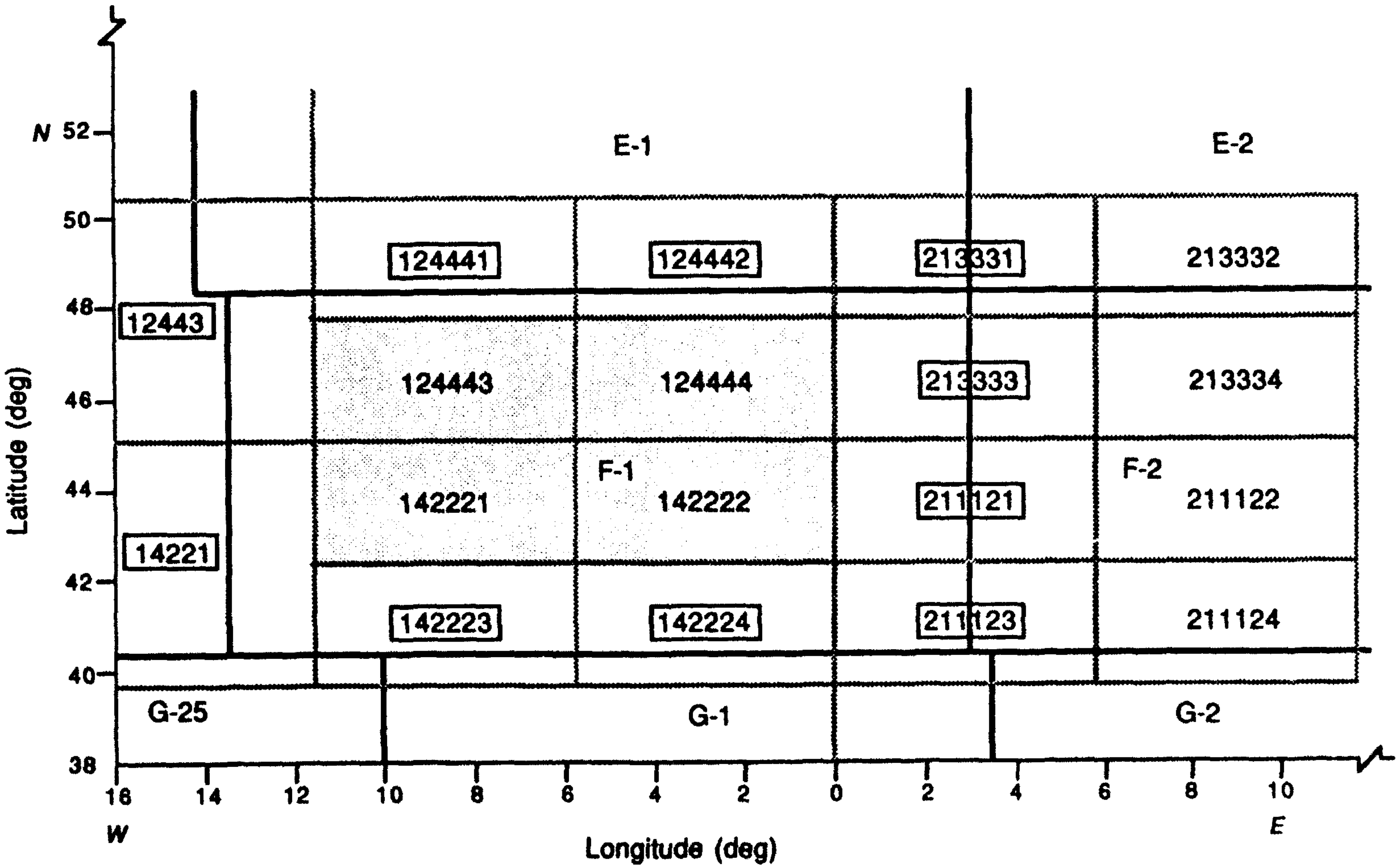

Figure 8. Characteristics of Adaptive Tiling of ONC F-1.

F-1 Tiling

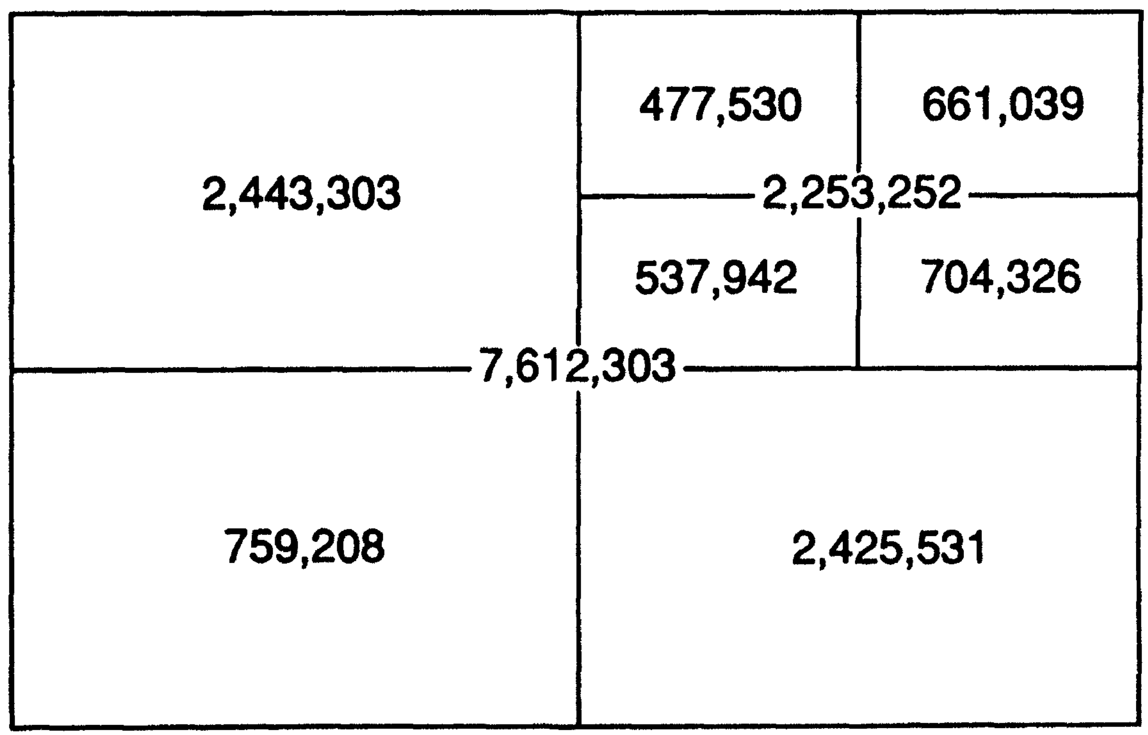

Figure 9. Comparison Between Whole and Quartered ONC G-18 Data Content (in Bytes).

ONC G-18 by Partition

2.2 ADAPTIVE TILING SUMMARY

Generally, adaptive tiling, by whatever measure, is a highly data-content-sensitive procedure that, through the data volume evaluation procedure, is closely coupled to all changes in data content and representation. Additionally, the final global tiling scheme will not be known until the last map sheet is automated and the final partition produced. This characteristic potentially presents an obstacle to geographic organization, or the arrangement of tiles, on the CDs. Adaptive tiling presumes that an optimal value not only exists but can be identified and executed in an implementation procedure.

Adaptive tiling, however, has many interesting properties that may have powerful implications for the examination, rather than the construction, of spatial databases. The ability of adaptive tiling, at a given level of measurement, to dynamically pass control among files or layers may permit new analyses of database morphology. Underlying database relationships, such as the impact of representation rules on database size or performance, may be exposed by extracting discrete data content measures obtainable through adaptive partitioning. Finally, and perhaps of most interest, is the potential for these discrete measures to permit investigations into the development of adaptive databases.

3.0 FIXED TILING

Fixed tiling is the other general approach that was considered for tiling the DCW. Fixed tiling is quite different from adaptive tiling in that it defines a set of fixed areas into which all data for that area is stored. Fixed tiles are partitions that, once defined, remain unaltered through subsequent processing. Since the resulting tile structure is, however, meant to partition space as a function of data content, constraints do exist. Therefore, the definition of tile size and shape in a fixed tiling scheme is best left until all influencing parameters are known and, insofar as possible, quantified.

As with adaptive tiling, implementing a fixed tiling scheme requires the development of a partitioning procedure as well as the determination of the unit of measure to be used to define the partition. In adaptive tiling, the determining unit of measure is data volume; in fixed tiling, it is spatial extent. It is possible to define an appropriate spatial extent for fixed tiling schemes because the various tiling constraints acting on the database interact in such a way that a range of data volumes with acceptable performance can be defined.

The fixed tiling approach is not required by definition to be either systematic or regular. (A more detailed discussion of the variations on fixed tiling is given in the Interim Tiling Design Study.) In the simplest case, a tiling scheme may be completely application driven and, therefore, "hard-wired" to support a single use. The partitioning of TIGER (Topologically Integrated Geographically Encoded Referencing System) files by county is an example of a fixed scheme driven by a specific application. Since the DCW is a general-purpose database for ad hoc use, an application-specific tiling scheme is inappropriate.

Systematic partitions along a longitude/latitude grid are a common and well-known form of spatial division. Examples of map products partitioned along longitude and latitude are numerous. They include the United States Geological Survey (USGS) topographic maps at scales of 1:250,000, 1:100,000, 1:63,360, 1:62,250, 1:50,000, 1:25,000, and 1:24,000; the Defense Mapping Agency's 1:500,000 Tactical Pilotage Charts (TPCs) , 1:250,000 Joint Operations Graphics, and 1:50,000 topographic series; and Digital Terrain Elevation Data (DTED) and Digital Feature Analysis Data (DFAD), both levels 1 and 3. The simplicity and wide availability of existing map products developed with systematic geographic partitioning make it this type of tiling attractive for use with the DCW. Therefore, the tiling scheme recommended for the DCW is tiling that is fixed and systematic.

3.1 FIXED TILING PROCEDURE

The procedure for generating a set of systematic partitions is operationally very simple. A fixed grid can be mathematically generated quickly and accurately for all or part of the globe. Figure 10 illustrates a systematic fixed tiling scheme drawn over the ONC map series outline. Figure 11 is an enlarged view of this same 5-degree-by-5-degree scheme for the area occupied by ONC F-1. Unlike adaptive tiling schemes, fixed tiling schemes do not require evaluations of data content. The tiling can begin as soon as the first layer within the map sheet is finished. Any map sheet may be automated and tiled without assessing the contents of incomplete map sections, even when additional map sheets must be prepared in order to finish incomplete tiles. This degree of independence between the tiling scheme and the incorporation of data is advantageous to a data producer, since the sequence of map sheet processing is not controlled by the tile procedure. Thus, in contrast to adaptive tiling, fixed tiling does not adversely affect production.

IndexLL.1b

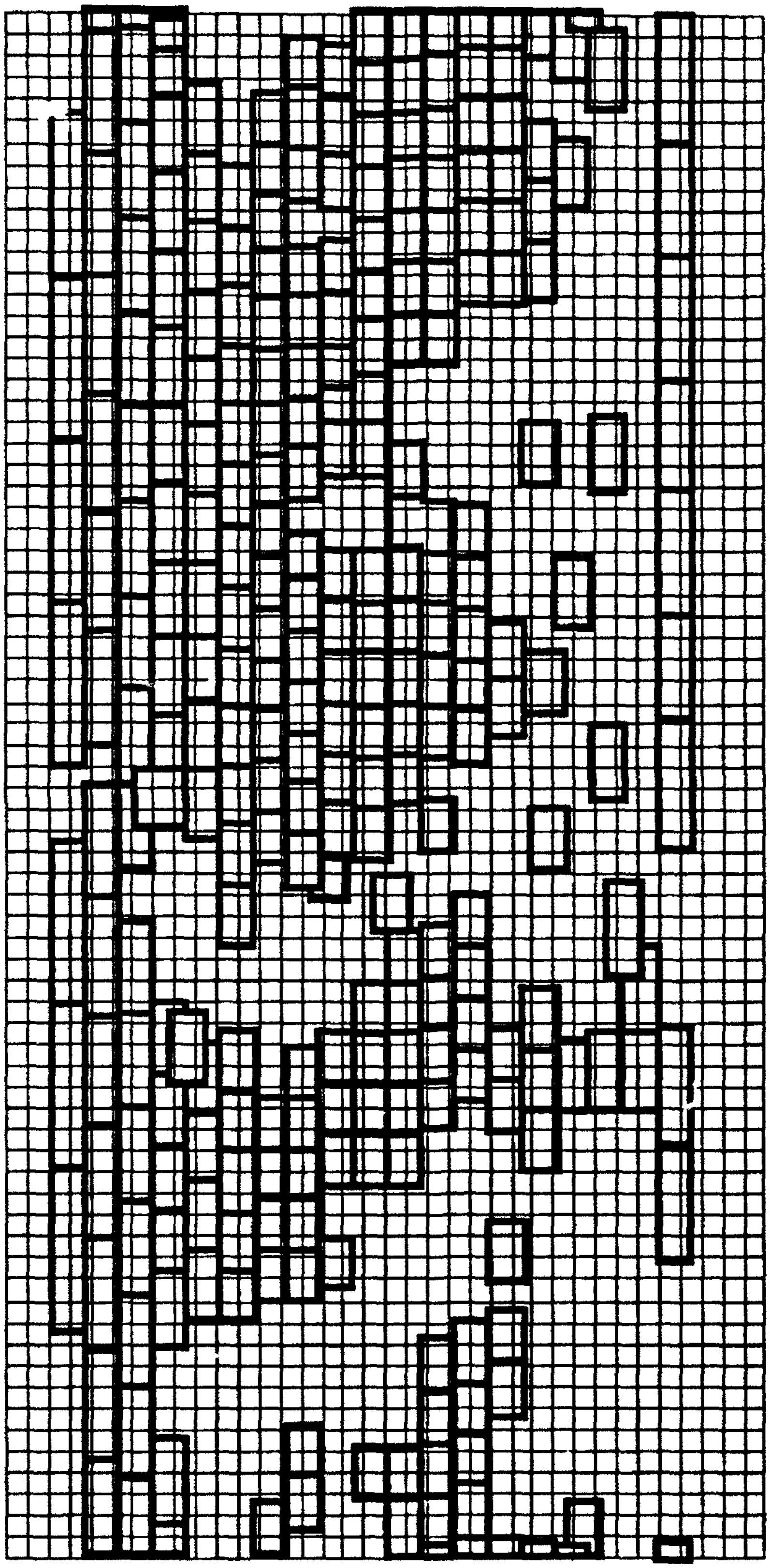

Figure 10. Global Fixed Tiling of the ONC Series.

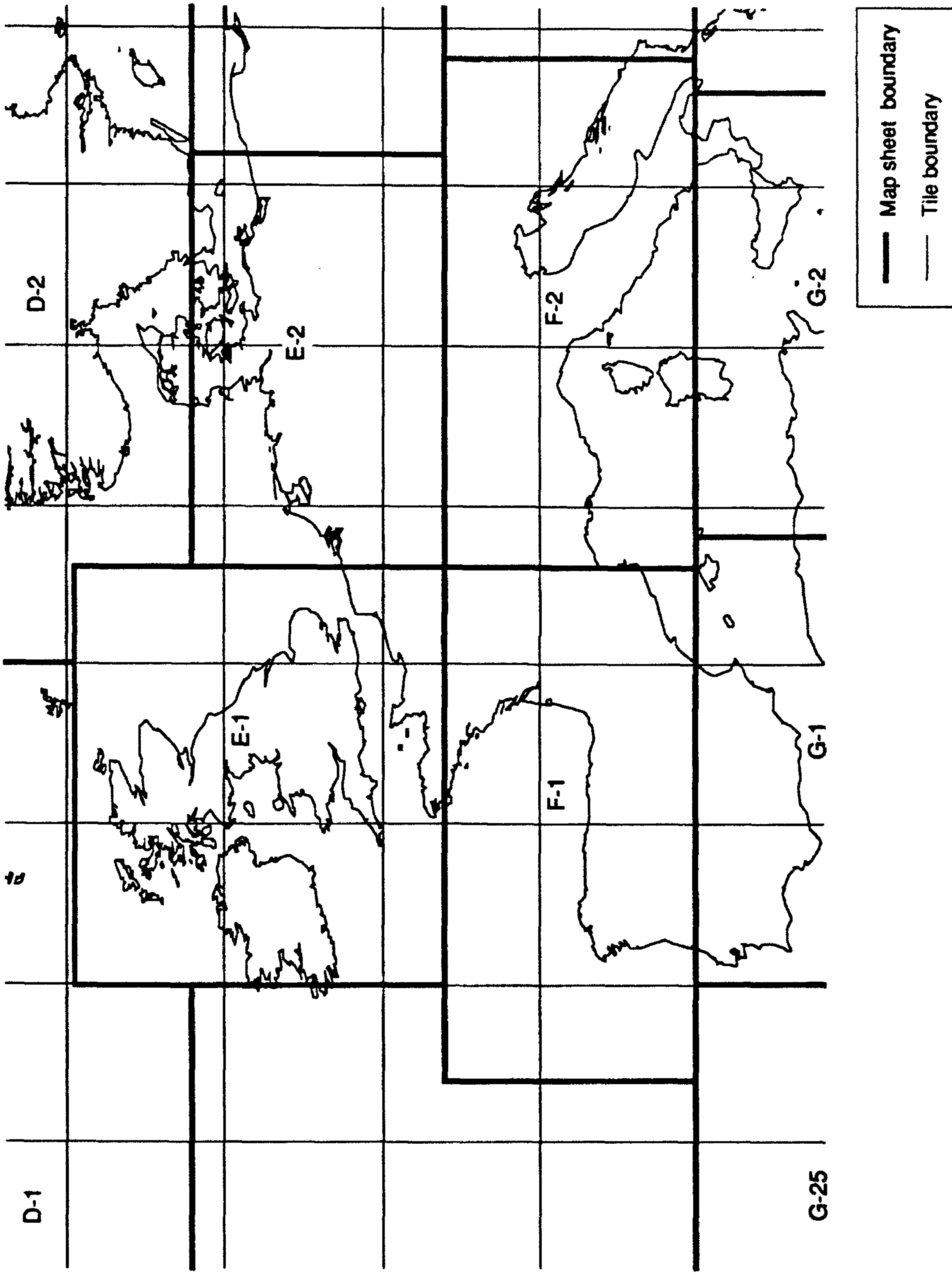

Figure 11. Fixed Tiles Near ONC F-1.

3.2 DETERMINING TILE SIZE

Whereas adaptive tiling sequentially evaluates the data content of each partition as it proceeds through the data, ensuring that no tile exceeds a maximum, a fixed tiling scheme defines partitions, or boundaries, prior to the addition of any data to the tile. A fixed tiling procedure provides a single tiling scheme for all layers. Fixed tiling, like adaptive tiling, determines a maximum value for data content to establish a tile size. But, unlike adaptive tiling, the determination of fixed tile size is a function not only of data content but also of the characteristics of the software and the secondary storage device. Each of these is discussed below.

The DCW display software processes queries in the following manner. The user first identifies a geographic area of interest, which may contain one or more tiles. The software then reads a tile data dictionary and returns to the user a list of the features contained within the selected area. The user selects features from the list; the software then accesses the files on the CD that make up the layers containing the features. If the user-selected area differs from the area covered by a tile or set of tiles, a tile spatial index returns the appropriate subset of features. Ongoing work on spatial indexing may permit tile size to increase while maintaining the same given level of performance.

The characteristics of the CD-ROM must also be factored into the tile size decision. Data retrieval from a CD-ROM involves first a seek, which moves the read head to the proper file position, and then a sequential read. By comparison with a magnetic fixed hard disk, CD-ROM seek times are relatively long, although read times are comparable. The long seek times act as a design incentive toward larger file size. Larger file size translates into larger tiles. The preliminary results of indexing studies indicated that a range of geographic extents, which correspond to a range of data content, provided an effectively linear response time against the CD-ROM. This means that within a certain range of geographic extents, query process times are linearly related to the number of features, for a given feature type. The tile size objective is then to stay within the data volumes that have been bracketed by those spatial extents.

For a given selected area, tiles that are smaller in volume than this linear response range of data volume will be penalized by the increased number of head seeks needed to access all the necessary files. That is, to draw a single layer from a tile requires seeks to all the files that make up that layer and then additional seeks to continue reading those files that exceed the one seek read size. In a situation where the tile size is above the optimal range, the number of files in a layer that exceeds this maximum size will be less than the total number of files in the layer. Continuing to read an overly large file will require fewer additional seeks than reading all the files for layers in overly small tiles. This is because the overly small tile situation requires multiple seeks to initiate access to all the tiles. Those tiles larger in extent than the upper range limit are expected to be penalized a lesser number of times to complete a sequential read from one or two overly large files. Therefore, there is a design incentive to reduce the absolute number of tiles and correspondingly increase the byte sizes of the files within each.

For the DCW, the data content range yielding a linear response was empirically determined to fall between a tile size of 3 degrees by 4 degrees and 4 degrees by 7 degrees using the drainage layer from ONC G-18. These data were in DCW format stored on the CD-ROM using ISO 9660. The reliability of this estimated data volume range may be judged by how closely ONC G-18 represents the overall data characteristics of the database. Figure 12 identifies the location of ONC G-18 on the estimated data distribution surface. The relative position of this sheet and the amount of data it contains indicate that it is representative of the midlatitude Northern Hemisphere, which is the area from which a large portion of DCW data will be obtained. ONC G-18 is at the 88th percentile for the source map sheets that will be used in the DCW, with respect to total estimated size. Only 35 sheets are estimated to contain more total data than ONC G-18. In addition, drainage, the layer chosen for testing, had the highest byte count of the 14 layers present for ONC G-18, indicating

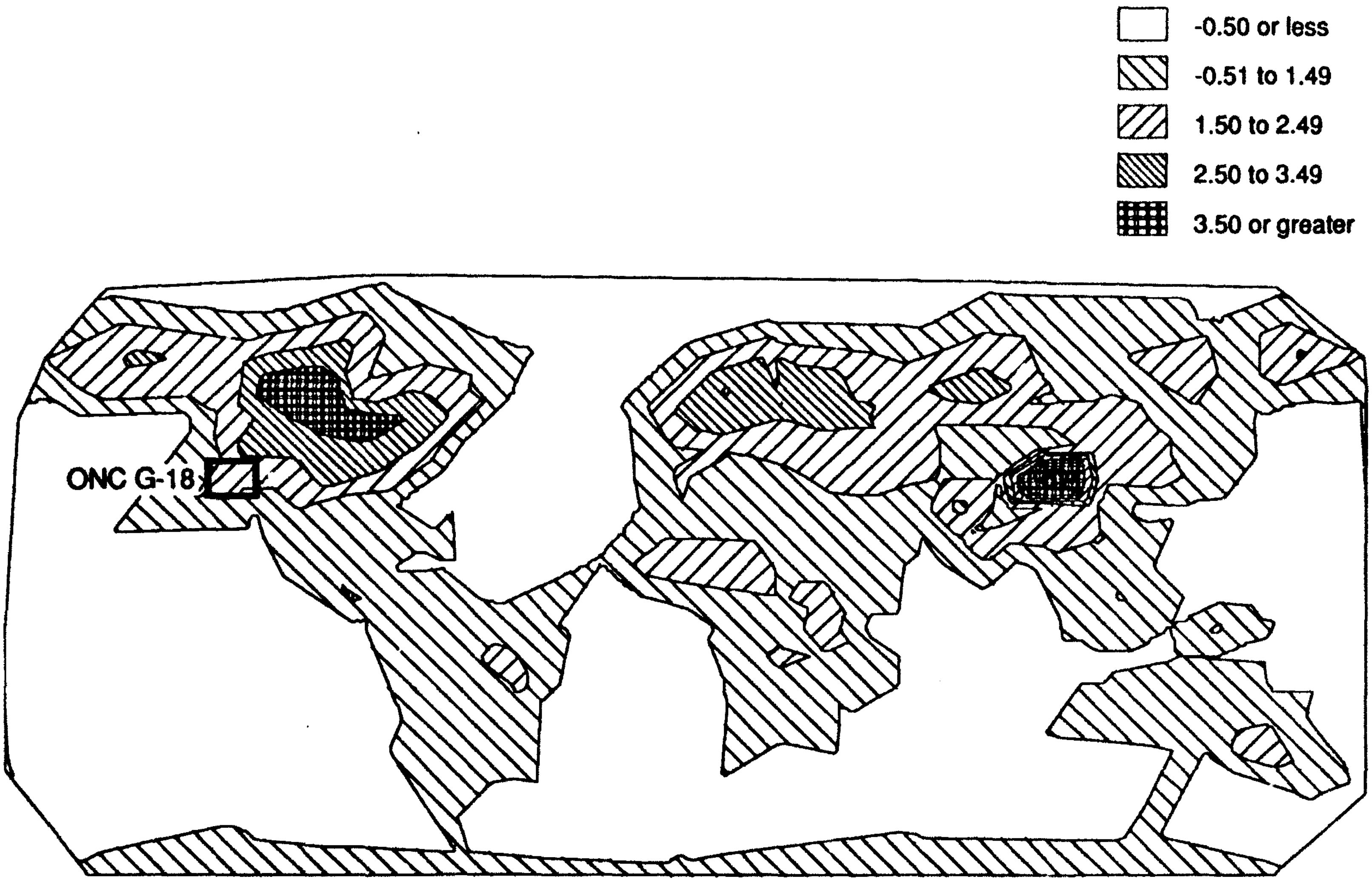

Figure 12. Position of ONC F-1 on the Estimated Data Distribution Surface.

TIN1.0G18b

that this layer's size represents the upper end of data content from a map sheet containing a significantly greater than average data volume.

The fixed tiling scheme has many advantages for the DCW database. It is simple, practical, easy to understand and create, is independent of the production process, and can provide the range of data volumes required by the CD-ROM. Using the test results from the indexing studies in order to empirically determine tile size, two fixed graticule-based tiling schemes were developed for evaluation in Prototype 4. These tiling schemes are both subsets of the existing World Geographic Referencing System (GEOREF). The first scheme is a 5-degree-by-5-degree tiling scheme (Figure 13); the second scheme consists of 3-degree-by-3-degree tiles. These two spatial extents correspond closely to the upper and lower file size range limits reported in the indexing studies.

The final tile size will be determined through measurements of map sheet data in DCW format evaluated against the CD. As such, the size will provide a general-purpose solution that reflects the interactions of the data structure, software functionality, secondary storage device, production constraints, geographic organization, the incorporation of other products, and user needs.

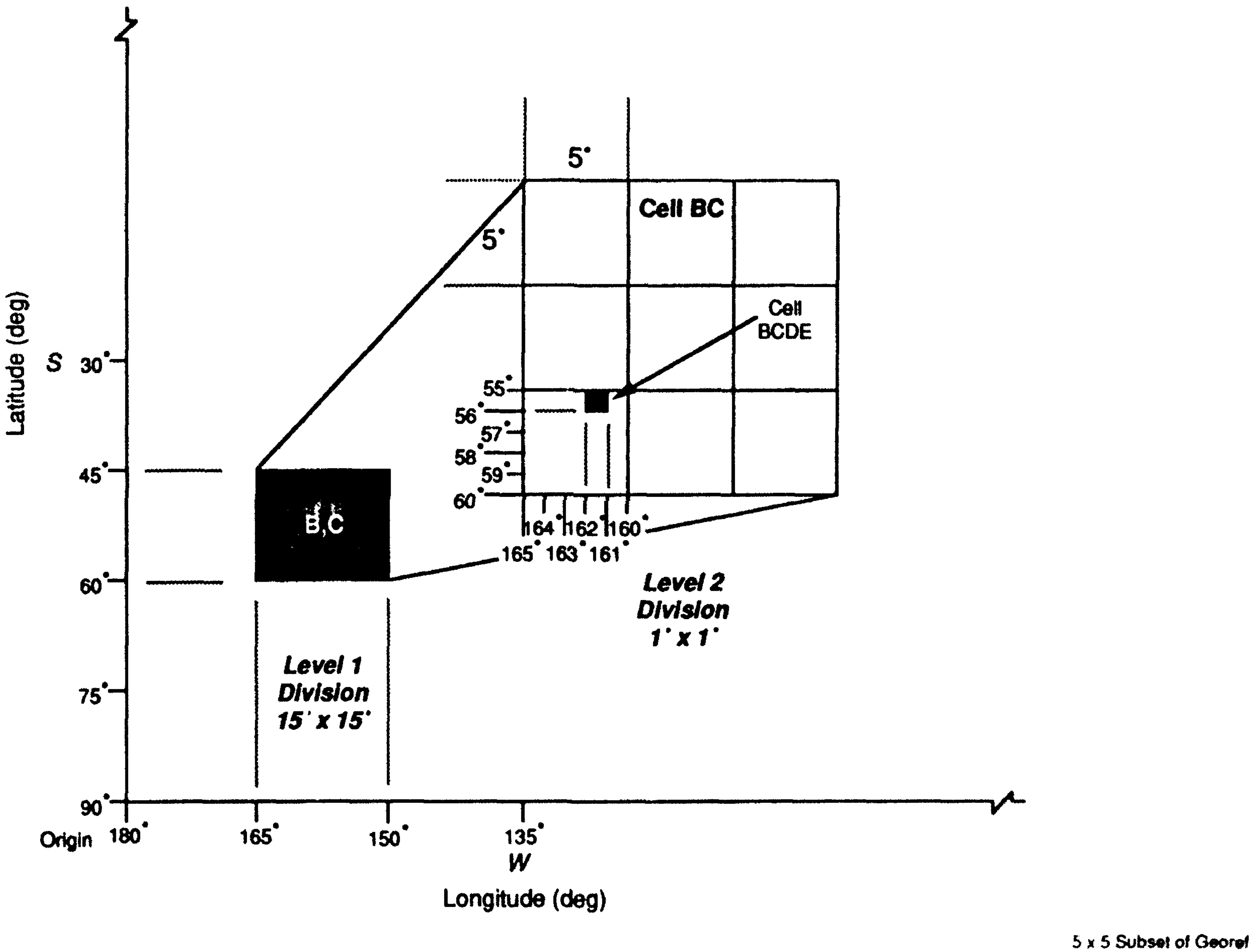

Figure 13. Fixed Tiling Subset of GEOREF.

4.0 TILING SCHEME IMPACTS ON CONVERSION AND CD-ROM CREATION

4.1 CONVERSION

Conversion is the process of translating the completed and tiled map sheets from the ARC/INFO format used in database production to DCW format. Stability in this process is particularly important in an adaptive tiling application since the measure of data volume occurs during the production stage. If a target value of data volume is determined and a tiling scheme is built around it, the conversion process must be predictable. If not, the actual byte counts may be changed during conversion and the level of accuracy attainable with adaptive tiling could be lost.

Table 3 presents file sizes before and after conversion for all layers present in ONC G-18, as well as the ratios of layer sizes between the two formats. These ratios vary between layers. The table demonstrates the volatility of these ratios and the resulting difficulty of applying the adaptive tiling procedure in that conversion environment.

Table 3. ONC G-18 Layer Sizes in Prototype 3 (All Layers).

Layer Name	ARC/INFO (1)	DCW (2)	Ratio (1)/(2)
AE	32,225	30,579	1.05
DN	1,140,097	1,176,518	0.97
DQ	6,743	6,875	0.98
MC	11,132	8,637	1.29
HS	190,188	230,952	0.82
HY	4,087,306	4,317,417	0.95
IS	3,502	3,878	0.90
LC	105,814	92,755	1.14
OF	2,348	2,480	0.95
PO	49,374	50,160	0.98
PP	367,830	312,592	1.18
RD	576,718	696,915	0.81
RR	251,090	308,565	0.81
UT	314,744	374,120	0.84
Total	7,139,111	7,612,443	0.94

One can assume, however, that a predictive model could be defined for an adaptive tiling scheme. Such a model could be defined after some trial conversion by testing a number of variable-sized layers and empirically defining a conversion factor to be included in the overall tiling procedure. Table 4, for example, presents the conversion ratios grouped by topologic type. While this procedure is not recommended, it does demonstrate greater stability in the conversion ratios when layers are arranged in this manner. This is largely due to the fact that each of the topologic groups has the same number of files for each layer.

Table 4. ONC G-18 Layer Sizes in Prototype 3 (Grouped by Topology).

Layer Name	ARC/INFO (1)	DCW (2)	Ratio (1)/(2)
AE	32,225	30,579	1.05
HS	190,188	230,952	0.82
IS	3,502	3,878	0.90
RD	576,718	696,915	0.81
RR	251,090	308,565	0.81
UT	314,744	374,120	0.84
DQ	6,743	6,875	0.98
OF	2,348	2,480	0.95
PO	49,374	50,160	0.98
DN	1,140,097	1,176,518	0.97
HY	4,087,306	4,317,417	0.95
GC	11,132	8,637	1.29
LC	105,814	92,755	1.14
PP	367,830	312,592	1.18

The variability in the data format ratios demonstrates that the conversion process can have considerable effect on data content byte counts, since these byte counts are influenced by representation rules and secondary storage standards. However, conversion has no impact on absolute spatial extents, whether the extents are generated by a fixed or adaptive procedure. That is, the coordinate values of tile boundaries are not affected by conversion although the data volumes of the files representing a particular spatial extent may be altered. Geographic extent is, therefore, very stable through the conversion process, while data volume is not. Therefore, the complex production process necessitated through an adaptive approach is further complicated by data volume unpredictibility in the conversion process.

4.2 CD-ROM CREATION

After the database has been produced, tiled, and converted to DCW format, it must be written to the CD-ROM medium. The tiling scheme and CD-ROM interact in two ways. First, the number of tiles present and the amount of data that each contains will directly influence storage space and overhead requirements. Secondly, the boundaries of each tile directly impact geographic organization. These interactions are discussed below.

4.2.1 Storage Requirements

During CD-ROM premastering, the database will increase in total size. This increase occurs because of the addition of data files supporting directory and indexing structures, and because of the padding that occurs since the ISO 9660 standard requires a minimum block size of 2048 bytes. Adaptive tiling schemes have the general property of producing fewer tiles for complete partitioning than a fixed scheme. Therefore, this method produces the smallest ISO 9660 padding, since a smaller number of files is created.

In the following example, storage increases that result from both additional directory structures and the ISO 9660 padding are evident. Table 5 presents the relationship between layer sizes for a whole map sheet and the sum of its quarters. The number of files present in the quartered version is four times as great as the whole version. In the quartered version, the total storage requirement is increased by 4 percent over the whole map sheet byte count. However, the distribution of the additional overhead is clearly skewed toward small content layers, with percentage overhead increases as layer size increases. Therefore, less overhead is incurred if the tile size is increased. Although the total number of tiles for an adaptive scheme will not be known until the database is completely partitioned, a data volume increase for each tile can be expected during the premastering process.

Table 5. Relationship Between Tile and Tile Subsets (ONC G-18 in DCW Format).

DCW Layer (1)	Whole Sheet Sum(bytes) (2)	Quartered Sheet Sum(bytes) (3)	Absolute Overhead (4)	Percent Overhead (4)/(2)
AE	30,579	37,152	6,573	21
DN	1,176,518	1,194,720	18,202	1
DG	6,875	14,870	7,995	116
GC	8,637	15,266	6,629	77
HS	230,952	237,922	6,970	3
HY	4,317,417	4,465,100	147,683	3
IS	3,878	11,264	7,386	191
LC	92,755	100,418	7,663	8
OF	2,480	8,624	6,144	248
PO	50,160	59,806	9,646	19
PP	312,592	322,422	9,830	3
RD	696,915	712,188	15,273	2
RR	308,565	317,202	8,637	3
UT	374,120	384,340	10,220	3
Total	7,612,303	7,881,294	268,991	4

Fixed tiling, in contrast, results in a known, although greater, number of tiles before the tiling procedure begins and allows more straightforward prediction of secondary storage needs. For example, if each tile were to contain all layers, then 188 files would be required to represent all features, irrespective of their actual data content. A 15-degree-by-15-degree tiling scheme would result in 288 tiles, or a total of 54,114 files for the DCW data. Using a 5-degree-by-5-degree tiling scheme would result in 2,592 tiles or 487,296 files. Likewise, the 3-degree-by-3-degree scheme would product 7,200 tiles and 1,353,600 files. Since speed on the CD-ROM is inversely related to the number of files present (that is, performance decreases as the number of files increases), there is a strong incentive to increase tile size.

Let us now examine more closely the padding overhead due to the effect of the ISO 9660 standard. The ISO 9660 standard allocates space on the CD-ROM in blocks of 2048 bytes, compared to a block size of 512 bytes in the production format. For this reason a large number of small files will incur more overhead than a small number of large files even though both have the same amount of data to manage. Total tile overhead is, therefore, a function of the feature representation, which will determine the number of files per tile, and the number of features, which will determine the size of each file. Again, from a tile design perspective, there is an incentive to reduce the number of tiles, which reduces the number of files, and increase the size of each.

The form of overhead, or padding, introduced by ISO 9660 will be greatest in tiles that contain little or no data. For example, assume that the 5-degree-by-5-degree tile scheme is used, yielding 2592 tiles. If no data will be made available for large water areas and these areas constitute 60 percent of the tiles, then approximately 1550 tiles will be empty. The empty tiles, carried as single faces, contain 10 files in VPF. Given the 2048-byte block factor in ISO 9660, 1550 tiles, each requiring 9 files of 2048 bytes each and one of 4096 bytes (22,528 bytes total), will occupy a total of 34.9 MB in the DCW. This volume will contain only the face identifying each tile as "ocean." (The 4096 bytes required for one file is due to densifying the tile boundaries for projection fidelity.)

If the 3-degree-by-3-degree scheme is used, 7200 tiles are required for global coverage. Since each of these tiles will have a smaller area, a greater number of them can be expected to lie in no-data areas. Assume that 65 percent, or 4680 tiles, are empty. This results in a total storage requirement for the ocean of 105 MB. In terms of storage overhead there are clear incentives to construct larger tiles even though the absolute amount of overhead is still very small compared to the total amount of available storage.

4.2.2 Geographic Organization

Geographic organization, the last procedure in which tiling has an impact on the database, refers to the process of regionalizing the DCW by distributing geographic sections of the globe onto different CD-ROMS. This final interaction of tiling with the CD-ROM occurs when the individual partitions are assembled into layered geographic regions that are then written to each CD-ROM. Figure 14 presents one of the recommended options of geographic organization. A characteristic of tiling can influence this process. It is desirable for the common boundaries to be along lines that properly demarcate the geographic divisions (e.g., continents) to be regionalized.

An adaptive tiling scheme will create a single demarcation of boundaries. However, the location of these boundaries with respect to the regional delineation will be determined only by the distribution of data and the depth of adaptive tile partitioning. Control of the regionalization is therefore taken away from the designer and becomes the arbitrary result of the adaptive tiling process.

In contrast, using fixed tile boundaries returns control of the geographic organization process to the partitioner. First, since only one tile scheme is produced, the tile boundaries of all layers within a library are explicitly in common. Second, their position with respect to the geographic areas of interest can be determined by generating a tile index and graphically overlaying it on the geographic areas to be regionalized. Those places where the tile boundaries do not satisfactorily represent the regional boundaries can be easily determined and corrected early in the tiling process.

These requirements of geographic organization present difficulties for tiling schemes that are not bounded on all sides. For example, ARC Digitized Raster Graphics (ADRG), considered a candidate tiling scheme, partitions the globe into bands of latitude that increase in interval width towards the equator. For ADRG to permit a variety of geographic organization options, it would need additional partitioning along selected lines of longitude.

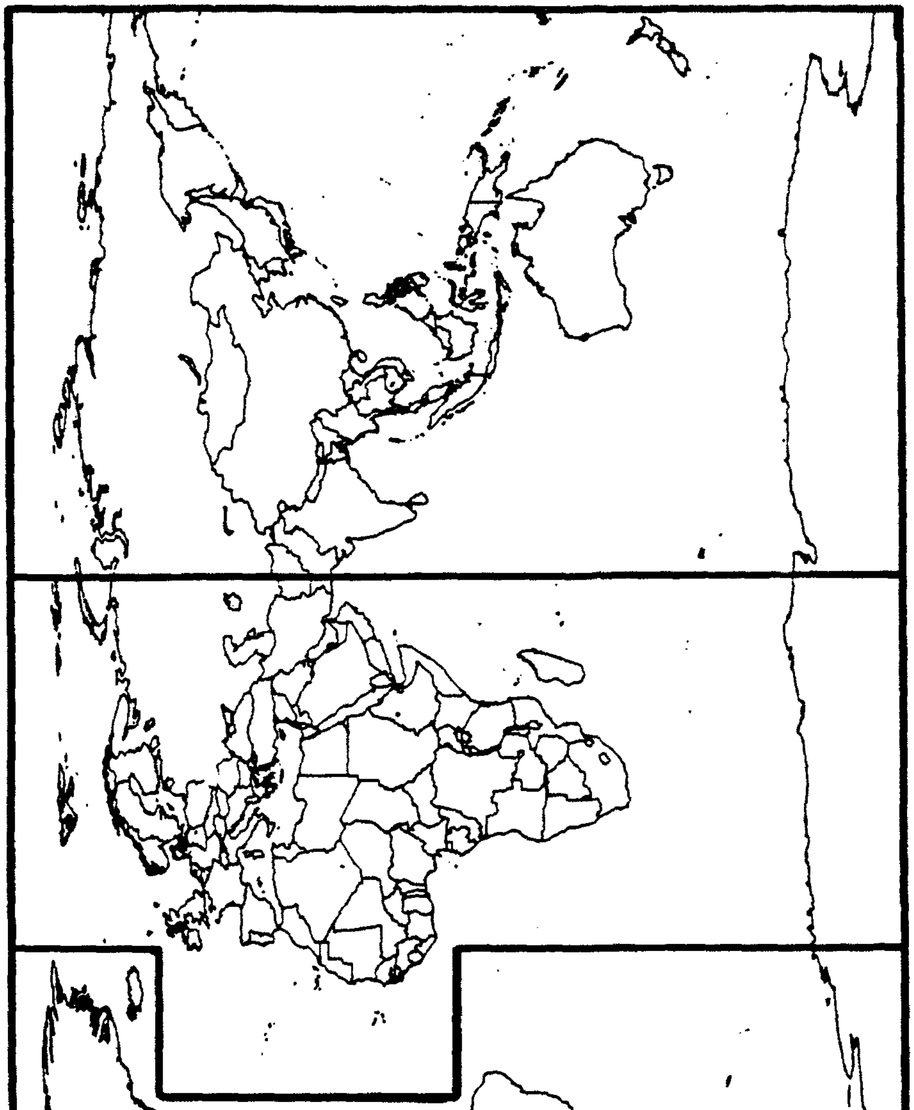

Figure 14. Geographic Organization that Minimizes the Number of Disks.

5.0 MAINTENANCE

Maintenance is the continuing process of modification to the database for purposes of adding new information or updating the current contents. Although a review of the interactions between tiling and database maintenance was not a requirement of this study, the impact of tiling on maintenance will be discussed briefly here. It is anticipated that maintenance activities will be extensive for the DCW database, because as many as 30 ONC sheets are missing roads data, 40 are missing contour information, and as many as 80 do not contain a vegetation layer.

Fixed tiling schemes with known spatial extents allow for the easy addition of new data, since they do not change as data content varies, although data additions can result in files and layers exceeding certain size thresholds, negatively impacting on CD-ROM performance. A summary table of the associations between tiles and map sheets can be generated with a single polygon overlay and maintained as a separate index. As they are required for maintenance, the tiles coincident with the map sheets to be amended can either be selected manually or accessed through a simple automated procedure.

As the DCW matures and becomes widely available, data users will likely become data producers. As these new data producers use a fixed tiling scheme, partitioning new data into spatial registration with the DCW will be easily accomplished.

6.0 APPLICABILITY TO OTHER PRODUCTS

The ability of a fixed tiling scheme of defined tile size to incorporate other products into the DCW can be assessed using the chosen tile size as a point of reference. Table 6 presents the scales, areas, and geographic coverage of four DMA map series that may undergo tiling at some future date. Each of these is a map series with ground coverage areas delineated along systematic divisions of longitude and latitude.

Table 6. Spatial Characteristics of DMA Map/Chart Series.

Map Series	Scale	Area (sq. in.)	Geographic Coverage (Long X Lat)
Topo	1:50K	407	15 min X 15 min
JOG	1:250K	503	2 deg X 1 deg
TPC	1:500K	1652	6 deg X 4 deg
ONC	1:1M	1714	14 deg X 8 deg

A fixed tile size such as that recommended for the DCW can be applied to larger-scale data. The tile size for these other map series can be estimated by comparing two characteristics of the larger-scale series to the ONC series: (1) the map scale and (2) the number of the larger-scale source sheets required to geographically cover the geographic area of a representative ONC sheet. That is, if the area covered by one ONC sheet is equal to 100 square degrees, then a map series that requires four sheets to cover that 100-square degree area would require a tile size that covers one-quarter of the ONC tile coverage. The assumption in these calculations is that map sheet feature density from series to series is roughly equal.

Table 7 records the number of sheets of each series required to cover an ONC, which normally has an 8-degree-by-14-degree coverage. A recommended tile size that is based upon the number of sheets required to geographically cover the ONC sheet is presented for each map/chart series.

The procedure used to create Table 7 consisted of dividing the number of sheets required to geographically cover the ONC extent into the area of the tile and then taking the square root of the result to yield the degree length for one side of a square tile for the larger-scale sheet. If the TPC information below is used and the ONC tile size is assumed to be 5 degrees by 5 degrees (one of the sizes to be evaluated in Prototype 4), the result is the square root of 25/4.7 or 2.3 degrees square. Since it is assumed that an even degree tiling procedure would improve maintenance for these larger-scale series, a range of 2 to 3 degrees is shown in Table 7. Similarly, for the 1:50,000 scale topographic series, the result is the square root of 25/1792 or 0.1 degree or 7.1 minutes square.

This estimation method presumes that the feature types and the number of feature occurrences of each is approximately common across scales. In addition, the assumption is made that a similar automation technology will be used on the larger-scale map sheets. Changes in capture resolution will result in changes to the number of coordinate pairs that are required to represent a feature and, thereby, changes to layer byte counts.

Table 7. Tile Size Recommendations Based on a 5 Degree by 5 Degree DCW Tile.

Map Series	Geographic Proportion of Sheets/ONC	Recommended Tile Size
ONC	1.0	5.0 deg square
TPC	4.7	2–3 deg square
JOG	56.0	0.5–1.0 deg square
Topo	1792.0	5–10 min square

7.0 SUMMARY

Table 8 presents a comparison of adaptive and fixed tiling in the context of the DCW. The first column identifies the tiling type. Each type is evaluated against the operation or device identified by the columns to the right. Since this table represents a procedural comparison only, neither tiling scheme specifies a particular tile size, or in the adaptive tile case, a measurement level. (See the appendix for a detailed discussion of measurement level). Assume that the tile size is a function of the total number of bytes in a layer, which is option 2 in the Appendix.

Table 8. Comparison of Adaptive and Fixed Tiling Schemes.

Type	Production	Additions	Geographic Organization	Storage	Other Products
Adaptive	—	—	—	0	—
Fixed	0	0	0	0	0

Beginning with impacts on production, adaptive tiling is a data-sensitive procedure that is considerably more difficult to implement in a production environment than fixed tiling because of its iterative requirements. With adaptive tiling, the tiling scheme cannot be known until all source material is automated and the last tile is produced. In addition, after the first map sheet is selected, the subsequent map sheets to be prepared are to some extent set by the tiling procedure. This inhibits the ability of the production procedure to respond to shifts in the areas of interest. A fixed tiling scheme imposes no constraints on either the order in which map sheets are to be prepared or the ability of the production process to respond to changes. In addition, a fixed tiling scheme allows any number of maps from any area of the world to be in the production stream. Fixed tiling can simply commence when a single layer from any map sheet is completed.

Next, data additions to an adaptively tiled database may require a re-evaluation of the tile size, depending on where the data is being added and in what amounts. Layer additions can be made to the database without retiling if the volume of new data does not exceed the volume of data used to define the original extent. If, however, new data does exceed the original tile limit, then that tile and all its layers will have to be partitioned. If the new data volume is less than the amount used to create the tile, then data additions can be made without re-evaluation. New data can be added to any tile in a fixed scheme without a data content evaluation. If the amount of data is significantly greater than the amount for which the original tile was designed, then performance for queries on this tile/layer combination may be reduced. In either fixed or adaptive tiling, however, the number of layers in a tile is ultimately limited only by the total number of bytes they collectively require, since no tile may contain more than the capacity of a single CD.

Third, adaptive tiling may impose significant constraints on geographic organization by generating tile boundaries inconsistent with the regional areas to be defined. A compounding problem will be the inability to predict the final outcome of the tiling scheme. In comparison, fixed tiles are defined before any data is added, so the tile boundaries can be evaluated for their effects on geographic organization before the tiling procedure begins. The ability to alter the tiling scheme to improve geographic organization is contingent upon other factors operating on the database, besides tile size. A very few large tiles, whether produced by an adaptive or fixed procedure, will be difficult to manipulate under either situation.

Fourth, the interaction of tiling and the storage media is basically determined by the number of tiles produced to partition the globe. The more tiles produced, the more files produced, along with the greater likelihood that small files may result. Any small file will receive a level of padding from

ISO 9660 that will reduce the level of storage efficiency. In general, an adaptive tiling scheme can be expected to produce a fewer number of tiles than a fixed scheme, resulting in fewer files, reduced overhead and higher performance. The amount of file overhead for tiling can be determined in advance with fixed tiling. Less tile overhead can be achieved through the use of larger tiles.

Finally, with respect to the applicability to other products, an adaptive scheme imposes the same level of procedural overhead and production interference with each new product to be tiled. This is again dependent upon whether data additions to a tile are in volumes greater than the original design volume. A fixed scheme provides an estimated tile size for other scale products. This is based upon the ratio of the developed tile scheme, source scale, and ground coverage to any other product source scale and ground coverage. This extendability to other products, plus the ability to tile individual sheets to an empirically measured performance value, are additional characteristics of fixed schemes.

8.0 CONCLUSIONS

In the context of the DCW project, adaptive tiling is a data-content-sensitive procedure that, in order to maintain accuracy, requires predictability in the post-tile processes of conversion and CD-ROM mastering. While some estimated measures are available, the true magnitude of all the influencing factors affecting adaptive tiling cannot be known until the partitioning procedure has been completed. Adaptive tiling is a dynamic procedure and does not have a known predictable structure usable by other groups involved in the DCW effort. Though acceptable for a stable database using partitioning as a selective optimization procedure, these characteristic hinder database construction efforts. Adaptive tiing is a scheme that is ill suited to a general-purpose database like the DCW's.

A single fixed tile scheme is a simple yet effective approach to tiling geographic databases. It is stable, easy to understand and create, and can be implemented independent of database construction and maintenance. Other project activities, including the development of geographic organization, software functionality, and data structure, can proceed uninhibited by the procedural requirements of tiling. These activities will aid in the determination of a final tile size, as will the evaluation of Prototype 4 and the results of the final indexing studies. Thus, a fixed tiling scheme allows tiling to accommodate information that is still being developed; and, all tile design issues considered, fixed tiling is the scheme recommended for the DCW database.

APPENDIX. DEFINITIONS OF DATA VOLUME PER TILE

In addition to developing an automated tiling procedure, an appropriate definition of tile data volume must be determined for adaptive tiling to be implemented. More precisely, a unit of measure for data density and a value for that measure are required to control the level of adaptive tiling. Figure 15 illustrates the four data volume definitions that have been developed for the DCW. First, tiles may be defined not to exceed a maximum number of bytes relative to the total number of bytes for all the data obtained from a map sheet; second, tiles may be defined not to exceed a maximum number of bytes relative to the largest DCW layer; third, tiling may be defined not to exceed a maximum number of bytes relative to the largest file within any layer; and finally, tiling may proceed among collections of layers that are grouped into common types of topology.

OPTION ONE: TOTAL BYTES PER TILE

Adaptive spatial partitioning based upon the maximum total byte count is a volume measure at the database level. That is, per unit data volume is determined from the sum of all files obtained from map sheet information, without regard to organization at the layer level. Figure 16 presents a hypothetical spatial database in cross-section to illustrate this concept. Since all of the tiles are meant to contain a variable, but less than maximum, value of data, their bounaaries will differ in distance along the x axis. Of the four levels of measurement to be discussed, the total number of bytes will be highest using this approach and, therefore, the required depth of partitioning will be greatest in order to yield a given byte value.

Tiling according to total data content is a measure that sums all cartographic and attibute data in the database into a single unit. Topologically integrated databases are, in effect, one layer and would, therefore, be adaptively tiled at this measurement level.

OPTION TWO: BYTES PER LARGEST LAYER PER TILE

The spatial partitioning of the database can be based upon a total byte count per layer measure, a relationship illustrated in Figure 15(2). The data to be evaluated is the sum of all files that make up each layer. The tiling procedure may then use the individual layer values to control the level of partitioning in two ways. First, a tiling scheme may be derived for each layer, which for the DCW would yield as many as 18 different schemes. A representation of the different tiling for the different layers is shown in Figure 17. The partitioning is at the layer level and can be manipulated at the layer, rather than the database, level. If this method is used, different tiling schemes within the DCW would need to be maintained as different VPF libraries. Because of the level of complexity introduced by separate layer tiling schemes and the requirement of a separate VPF library to manage each, this separate layer/tile scheme option is not recommended.

However, the entire database can be tiled into a single scheme using the value of the largest layer for a map sheet as the control. That is, an area is defined from a base of the largest layer which, because the layer contains less data than the total map sheet for the same area, will yield larger tiles. Each of these tiles is then used to partition the remaining layers. Figure 18 presents the same data as Figure 17, except that in this example the database is disaggregated and the layers are placed on the same base to allow for an easy comparison of absolute size. Wherever a layer has the greatest amount of data, identified by the bold line along the top edge, it will rise above the other layers and provide the control for tiling. Wherever another layer is larger, it will take on the control function. In the example, UT is the first layer exercising control; then DN does, followed by RD, and so on. In this manner, the control of the database partitioning measure passes among many layers, introducing the effects of feature representation as well as the number of feature occurrences. This measurement method may allow tiling to influence the database in a manner that is difficult to assess.

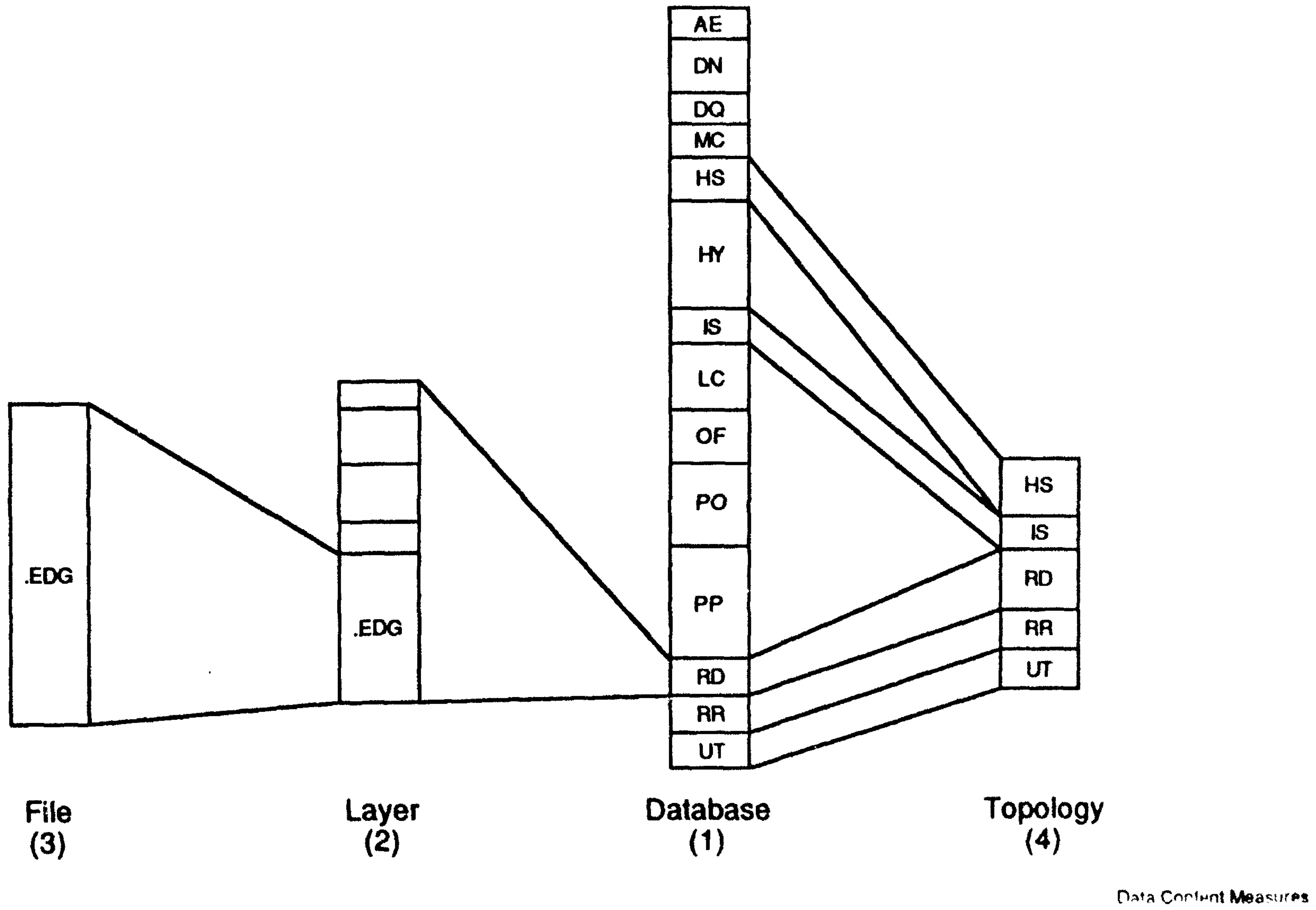

Figure 15. Definitions of Data Volume Used in Adaptive Tiling.

Fig. FF-Global Distrib.

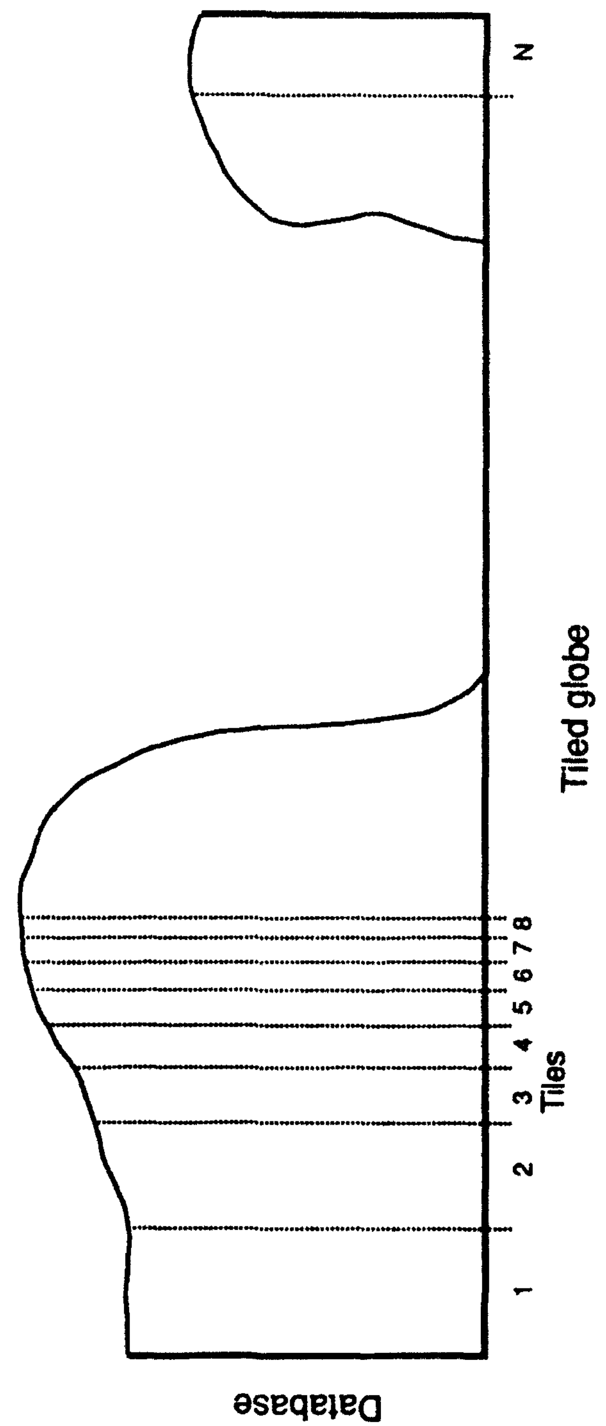

Figure 16. Cross-Section of a Hypothetical Spatial Database.

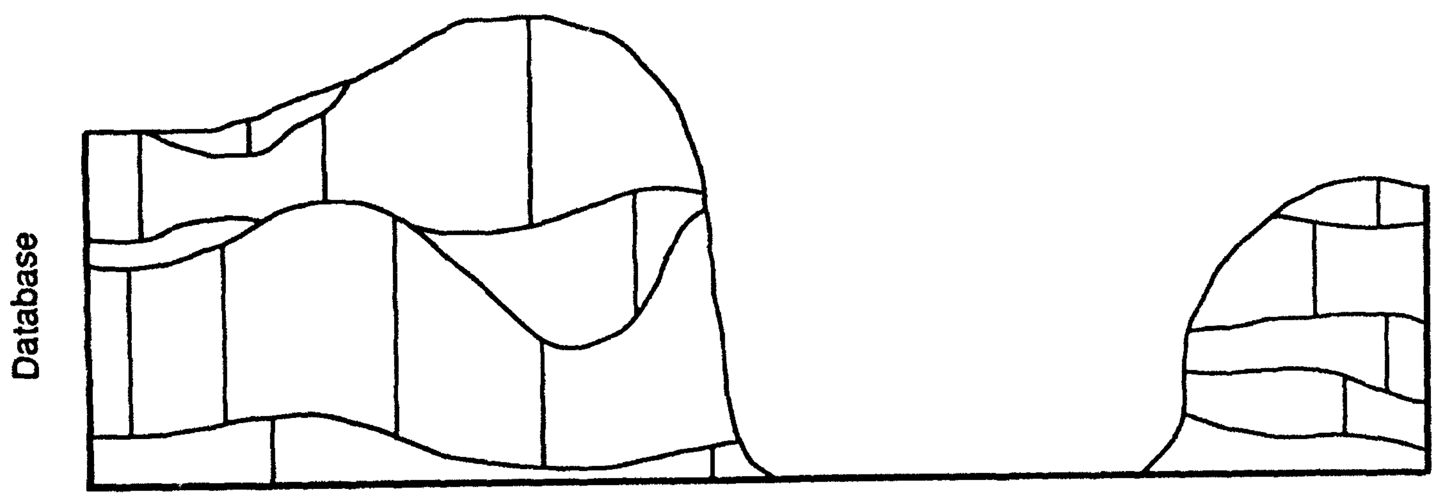

Figure 17. Hypothetical Spatial Database Tiled at Each Layer.

Fig. FF(a) Global Dist.

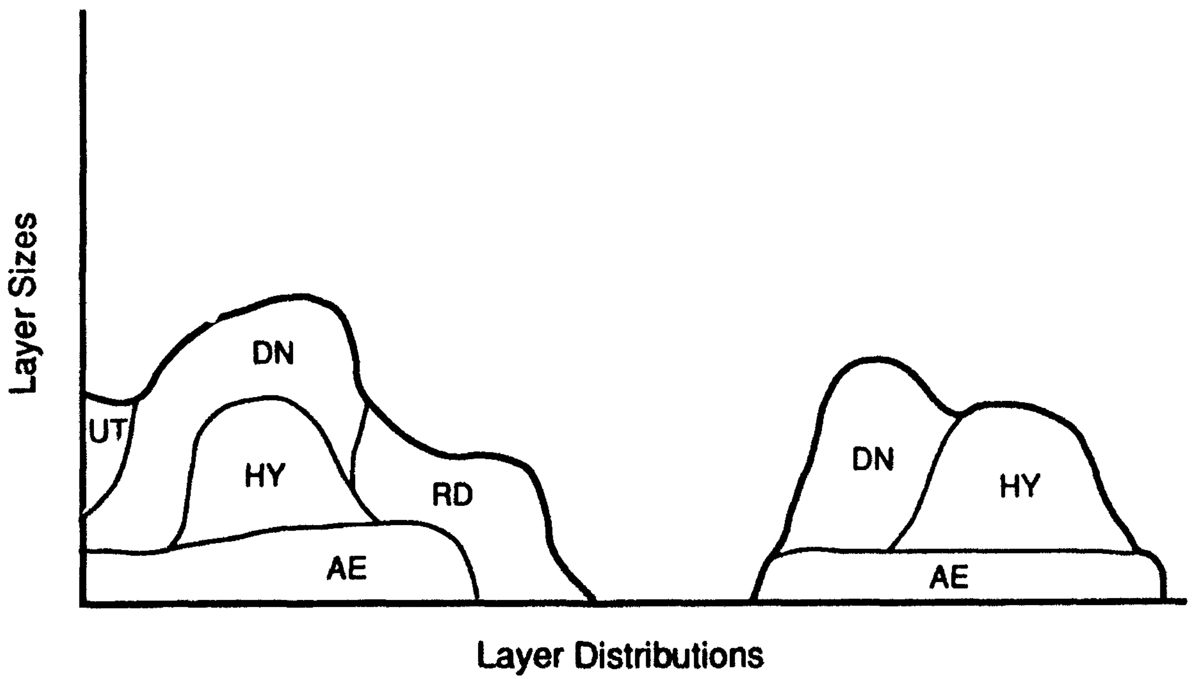

Figure 18. Hypothetical Spatial Database Tiled By Maximum Layer Measure.

Tile Largest Layer

Table 9 illustrates the byte counts by layer for the map sheets produced for Prototype 3. Notice that because of high spatial heterogeneity, the layer controlling the adaptive tiling would change from drainage in ONC E-18 to hypsography in ONC G-18 and back to drainage in ONC N-13 and JNC 120.

The net effect of this procedure is the generation of a single tiling scheme that is used to partition all layers but is derived from the largest. The use of a single tiling scheme to partition all layers has several advantages. First, part of the VPF implementation is the creation of a tile data dictionary, the purpose of which is to return to the user the data contents of the selected area. This is most easily performed by reading a single file referencing a single tiling scheme. A single tiling scheme also provides for a band-interleaved-like placement of layers on the CD, a desirable characteristic, given that data users are more likely to want more information about a given place than the same thematic information for other places. Finally, by defining a tile from a base of the largest layer on the map sheet, individual layers can be added to the tile, so long as none exceeds the size in bytes of the layer used for the original tile construction. That is, if the data volume measure were, for example, 1 MB, then additional layers each with up to 1 MB of data for that tile extent could be added without retiling. The maximum data content of a tile is therefore not limited by the content of any single layer but by the total content of all layers and cannot exceed the capacity of a CD.

Table 9. Data Variability (in Bytes by Layer) for Prototype 3 Sheets in DCW Format.

Topologic Group	No. of Files	Layer	ONC E-18	ONC G-18	ONC N-13	JNC 120
1	4	AE	4,447	30,579	2,943	3,992
5	18	DN	11,613,977	1,176,518	387,249	557,839
3	10	DQ	7,515	6,875	6,715	5,351
6	6	GC	4,323	8,637	16,500	10,381
2	6	HS	0	230,952	0	313,774
5	18	HY	2,966,283	4,317,417	328,227	11,019
2	6	IS	8,292	3,878	3,764	8,620
7	14	LC	266,611	92,755	240085	0
4	16	OF	0	2,480	243,205	0
2	6	PH	33,310	0	0	0
4	16	PO	276,280	50,160	297,540	434,428
7	14	PP	35,692	312,592	8,453	0
2	6	RD	150,021	696,915	131,829	0
2	6	RR	43,703	308,565	5,781	0
2	6	UT	51,160	374,120	0	0
Totals			15,439,402	7,617,531	1,677,379	1,350,492

Since a layer is a subset of the entire database, each layer total byte count will be smaller than the total for the database as a whole. Adaptive tiling using this measurement method, therefore, should result in tiles that are larger in extent and fewer in number than the tiles produced under the option 1 measurement method.

OPTION THREE: BYTES PER LARGEST FILE PER TILE

A third measure of data volume identified for adaptive tiling is at the file-within-layer level. This measure is identical in concept to option 2 above, but at a finer level of resolution than the layer. Figures 15(2) and 15(3) illustrate the relationship between layers and files. Depending upon the primitives required to represent the features present, each DCW layer can contain from 4 to 18 files. Since the individual file is a subset of the layer, the values at this measurement level will be

smaller than either the database or layer level and, therefore, result in the smallest number of tiles with the largest extents.

While at the layer level control of the tiling procedure may move among several layers, each of which is maximum at a different place, control of the tiling procedure at the file measurement level will move among different files from different layers. That is, for a given map sheet the selection of the largest file within a layer to drive tiling implies that the procedure does not distinguish between representations of cartography and attribution. Therefore, where cartography, according to file size, is the dominant map element, the data volume limit to partitioning is likely to be assigned from coordinate representations of place. Conversely, in places having large amounts of attribution and simple cartography, spatial partitioning could be driven by the thematic characteristics of place.

OPTION FOUR: BYTES FOR GROUPS OF LAYERS WITH COMMON TOPOLOGY

The last approach derives the partitioning measure from sets of layers that are grouped into categories having the same type of topology, a relationship illustrated in Figure 15(4). That is, all layers comprised of just edges are combined into a group, all layers that are just entity points are combined into a group, and so on. This effectively assembles layers into groups having the same representation rules and the same number of files.

Table 10 presents the toplogic type and number of files, by layer, for Prototypes 3 and 4. Notice that changes in database representation have taken place since Prototype 3, eliminating layers comprising integrated faces and edges. Also, although Prototype 4 now has four additional layers, they are all combinations of topology that existed in Prototype 3.

Table 10. DCW Layers Grouped by Common Topology.

Topologic Group	Layer(s)
Prototype 3	
Entity Points	AE
Edges	RR,RD,UT,IS,HS,OF
Faces	DQ
Edges and Faces	PO
Faces and Entity Points	PP,LC
Edges and Entity Points	GC
Edges, Faces and Entity Points	DN,HY
Prototype 4	
Entity Points	AE,DS
Edges	RR,RD,UT,PH
Faces	VG,DQ
Edges and Faces	—
Faces and Entity Points	PP,LC,LM
Edges and Entity Points	GC,HS,OF,TS
Edges, Faces and Entity Points	DN,HY,PO

Tiling at this common topology level presumes that features with the same representation will have common characteristics that can be exploited as a tiling design tool. Using this method of measure, those layers that are members of the same representation structure are compared to find the largest. Partitioning then occurs in a manner identical to the maximum layer measure explained in the

section on option two. Tiling by this measure should yield as many tiling schemes and VPF libraries as topologic groups, which for the current DCW is six.

EVALUATION OF MAXIMUM DATA VOLUME OPTIONS

The first measurement option, the database level, is intuitive and is the measure presumed in the literature on adaptive tiling. A significant limitation, however, is that once a scheme is developed under this measure, any additions to the database will require it to be re-evaluated and perhaps a portion of it retiled, if partitions into which data is being added exceed the specified maximum. Therefore, measurement at this level is best considered only for a database that is static. With regards to the display software, a tile volume measured at this level presumes that operations on the database are in whole tile units, which may be appropriate only for a topologically integrated database.

The second measurement option for adaptive tiling, control by the largest layer, is more attractive than the database approach for several reasons. First, sizing a tile in response to the largest layer makes the addition to the database of any smaller layer easy. For instance, if drainage is the largest layer in an area, then additions of other layers in that area up to the byte size of drainage are allowed without re-evaluation of the affected tiles. Also, tiling at the layer level of measurement is appropriate, since this is the thematic level at which users, through the display software, interact with the database. The layer is the atomic unit of the DCW, since a layer cannot be decomposed into its set of files and still retain the thematic meaning. A point of concern about using this measurement level is the extent to which tiling influences the database as control is passed among layers.

The third option, measurement at the file within layer level, is a method that offers the potential for performance improvements on the CD if file sizes are limited to those that allow the most efficient access from the disk. Additionally, tiling at this measurement level will produce the smallest number of tiles and, therefore, tiles of the largest size for a given amount of data and a given maximum per-tile data volume. Again, though, there is concern about generalization, caused, in this case, by the control of tiling being passed among many different files.

The fourth option, tile measurement obtained from a collection of layers having common feature topology, may provide a greater level of tile content predictability in subsequent processing. Since the groups are defined to contain features constructed using the same representation rules and, therefore, the same number and type of files, only the absolute feature counts will influence the tiling volume rather than combinations of representation rules and a relative mix of cartography and attribution.